

STIC-Biotech/ChemLib

97013

**From:** Slobodyansky, Elizabeth  
**Sent:** Thursday, June 19, 2003 12:24 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** 09/889,609

Please search for case 09/889,609:



SEQ ID NOS: 1 and 8-10 against commercial ~~and interference~~ databases.

Thank you.

*Elizabeth Slobodyansky, PhD*

Primary Examiner

Art Unit 1652  
CM1 10D11  
703-306-3222

**mail box 10D01**

Mary Jane Ruhl  
Tech. Info. Specialist, STIC  
TC-1600  
CM-1, Room 6A-06  
Phone: 605-1155

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model.

Run on: June 24, 2003, 06:24:52 ; Search time 2668 Seconds  
(without alignments)  
11158.984 Million cell updates/sec

Title: US-09-889-609B-9

Perfect score: 1023  
Sequence: 1 atgtgtctcagttatgcat.....atcagtcgtttctgttaa 1023

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number\_of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl:\*
- 1: gb\_ba:\*
  - 2: gb\_btg:\*
  - 3: gb\_in:\*
  - 4: gb\_on:\*
  - 5: gb\_ov:\*
  - 6: gb\_pat:\*
  - 7: gb\_ph:\*
  - 8: gb\_pl:\*
  - 9: gb\_pr:\*
  - 10: gb\_ro:\*
  - 11: gb\_sts:\*
  - 12: gb\_sy:\*
  - 13: gb\_un:\*
  - 14: gb\_vi:\*
  - 15: em\_ba:\*
  - 16: em\_fun:\*
  - 17: em\_hum:\*
  - 18: em\_in:\*
  - 19: em\_mu:\*
  - 20: em\_on:\*
  - 21: em\_or:\*
  - 22: em\_ov:\*
  - 23: em\_pat:\*
  - 24: em\_ph:\*
  - 25: em\_pl:\*
  - 26: em\_ro:\*
  - 27: em\_sts:\*
  - 28: em\_un:\*
  - 29: em\_vi:\*
  - 30: em\_htg\_hum:\*
  - 31: em\_htg\_inv:\*
  - 32: em\_htg\_other:\*
  - 33: em\_htg\_mus:\*
  - 34: em\_htg\_pln:\*
  - 35: em\_htg\_rod:\*
  - 36: em\_htg\_mam:\*
  - 37: em\_htg\_vrt:\*
  - 38: em\_sy:\*
  - 39: em\_htgo\_hum:\*
  - 40: em\_htgo\_mus:\*
  - 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1023	100.0	1023	9	AY034081	AY034081 Homo sapi
2	1023	100.0	1166	9	AF169974	AF169974 Homo sapi
3	1023	100.0	1770	6	AX224418	AX224418 Sequence
4	1023	100.0	2477	9	AK023169	AK023169 Homo sapi
5	1021.4	99.8	1336	6	AX259575	AX259575 Sequence
6	1020	99.7	1020	6	AX224420	AX224420 Sequence
7	867	84.8	2264	9	HSM805450	AL834378 Homo sapi
8	816.6	79.8	1197	10	AF148321	AF148321 Mus muscu
9	816.6	79.8	1411	10	BC011164	BC011164 Mus muscu
10	221	21.6	3807	9	BC019090	BC019090 Homo sapi
11	221	21.6	4107	9	AB037822	AB037822 Homo sapi
12	221	21.6	4140	9	AK026565	AK026565 Homo sapi
13	221	21.6	123454	9	HSEC17A99	AL450226 Homo sapi
14	221	21.6	221909	2	AC021705	AC021705 Homo sapi
15	181.6	17.8	184151	10	AL604066	AL604066 Mouse DNA
16	181.6	17.8	205910	2	AC044847	AC044847 Mus muscu
17	180	17.6	3379	10	BC031531	BC031531 Mus muscu
18	176.2	17.2	101814	2	AC122087	AC122087 Rattus no
19	157.6	15.4	40005	8	SPCC320	AL022245 S.pombe c
20	157.6	15.4	42689	8	SPCC330	AL031603 S.pombe c
21	157.2	15.4	6525	8	SCYKL218C	Z28218 S.cerevisia
22	157.2	15.4	24577	8	SC6ORF	X75951 S.cerevisia
23	142	13.9	295350	1	AP004826	AP004826 Staphyloc
24	137	13.4	137	11	G32814	G32814 A009527 Hum
25	131.4	12.8	298050	1	AP003189	AP003189 Clostridi
26	128.4	12.6	291050	1	AP000982	AP000982 Sulfolobu
27	126.2	12.3	16931	1	AE001716	AE001716 Thermotog
28	126	12.3	303750	1	AP003133	AP003133 Staphyloc
29	126	12.3	346900	1	AP003362	AP003362 Staphyloc
30	114	11.1	348050	1	AP003581	AP003581 Nostoc sp
31	111.4	10.9	303040	1	AP000991	AP000991 Thermopla
32	110.4	10.8	9990	1	AE010646	AE010646 Fusobacte
33	110.2	10.8	157609	2	AC095280	AC095280 Rattus no
34	105	10.3	268200	1	AP000061	AP000061 Aeropyrum
35	105	10.3	317511	1	CJ11168X3	AL139076 Campyloba
36	104	10.2	1631	3	AY052059	AY052059 Drosophil
37	100.2	9.8	35437	1	SCF43A	AL096837 Streptomy
38	98.4	9.6	166050	1	AL646085	AL646085 Ralstonia
39	97.8	9.6	14449	1	AE004696	AE004696 Pseudomon
40	96.6	9.4	1779	8	AF096281	AF096281 Arabidops
41	96.6	9.4	1779	8	AF177212	AF177212 Arabidops
42	96.6	9.4	2056	8	AY065037	AY065037 Arabidops
43	94.4	9.2	10805	1	AE011664	AE011664 Xanthomon
44	92.8	9.1	8044	8	SPBC1677	AL035581 S.pombe c
45	89.4	8.7	10881	1	U32757	U32757 Haemophilus

ALIGNMENTS

RESULT 1	AY034081	AY034081	1023 bp	mrna	linear	PRI 01-JUN-2002
LOCUS	AY034081	Homo sapiens	serine racemase mRNA, complete cds.			
DEFINITION	AY034081	Homo sapiens	serine racemase mRNA, complete cds.			
ACCESSION	AY034081					
VERSION	AY034081.1	GI:21307620				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
TITLE	1 (bases 1 to 1023)					
	Xia,M., Liu,Y. and Connolly,T.M.					
	Molecular cloning and expression of human serine racemase from					
	human NT2N'cells					

601	Db	AAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAATATGCAGATGACTGCTACCAAGTCC	660
661	Qy	AAAGCTGAAGGGGAACTGATGCCCAATCTTTATCTCTCCAGAAACCATAGCAGATGGTGTCT	720
661	Db	AAAGCTGAAGGGGAACTGATGCCCAATCTTTATCTCTCCAGAAACCATAGCAGATGGTGTCT	720
721	Qy	AAATCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGAGACCTTGTGGATGATATCTTTC	780
721	Db	AAATCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGAGACCTTGTGGATGATATCTTTC	780
781	Qy	ACTGTCCACAGAGATCAAAATTAAGTGTGCAACCACAGCTGGTGTGGGAGAGGATGAAACCTA	840
781	Db	ACTGTCCACAGAGATCAAAATTAAGTGTGCAACCACAGCTGGTGTGGGAGAGGATGAAACCTA	840
841	Qy	CTCATTTGAACCTACAGCTGGTGTGAGTGGGTGCTGTCTGTCTCAACATTTTCAAACT	900
841	Db	CTCATTTGAACCTACAGCTGGTGTGAGTGGGTGCTGTCTGTCTCAACATTTTCAAACT	900
901	Qy	GTTTCCCCCAGAGTGAAGAACATTTCTATTGTGCTCAAGTGGTGGAAATGTAGACTTAAACC	960
901	Db	GTTTCCCCCAGAGTGAAGAACATTTCTATTGTGCTCAAGTGGTGGAAATGTAGACTTAAACC	960
961	Qy	TCTTCATTAACCTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTCTTCTGTT	1020
961	Db	TCTTCATTAACCTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTCTTCTGTT	1020
1021	Qy	TAA 1023	
1021	Db	TAA 1023	
RESULT 2			
AF169974			
LOCUS		AF169974	
DEFINITION		Homo sapiens serine racemase mRNA, complete cds.	
ACCESSION		AF169974	
VERSION		AF169974.1	
KEYWORDS		GI:11034784	
SOURCE		Homo sapiens.	
ORGANISM		Homo sapiens.	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		De Miranda, J., Santoro, A., Engelder, S. and Wolosker, H.	
JOURNAL		Human serine racemase: molecular cloning, genomic organization and	
MEDLINE		functional analysis	
PUBMED		Gene 256 (1-2), 183-188 (2000)	
REFERENCE		20510003	
AUTHORS		2 (bases 1 to 1166)	
TITLE		De Miranda, J., Santoro, A.H., Engelder, S. and Wolosker, H.	
JOURNAL		Submitted (18-JUL-1999) Departamento de Bioquímica Médica, Rio de	
REFERENCE		Janeiro, Av. Brigadeiro Trompowski, s./n, Rio de Janeiro 21491-590,	
TITLE		Brazil	
JOURNAL			
FEATURES			
source		Location/Qualifiers	
		1..1166	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/tissue_type="brain"	
		17..1039	
		/function="catalyzes L- to D-serine racemization"	
		/codon_start=1	
		/product="serine racemase"	
		/protein_id="AAG27081.1"	
		/db_xref="GI:11034784"	
		/translation="MCAQYCSFADVEKAHINIRDS IHLTPVLTSILNQLTGRNLFP	
		KELFQKTSRKINALNARSLVPDALERKPAVVTSHSSNGHQAULTYAKLEGIA	
		YVQPTAPDCKKIAIQAYAS IYCEPSDESRENKAVRTEEGINMHVNPQAPVI	
		AGQGTALIEVNLQVPLADVPVGGGGMLAGIATVRLKPSVKYVAEIPSNADQV	
		QSKLKGKLPNLYPPTIADGVKSS IGLNTWPI IRLVDDIIFTVTDEIKCATOLWE	
		RMKLLIETAGVGVAALVSOHFOTVSPVKNIC IIVSGNVDLTSITVDEIVKQARPA	
CDS			



```
Db 429 TACGGACGCTCAATTTGTATATCTGTGAACCTAGTGTGATGAGTCCAGAGAAAATGTTGCAAAA 488
Qy 421 AGACTTACAGAAAGAACAGAGGATCATGTGTATCATCCCAACAGGAGCTGCGAGTGATA 480
Db 489 AGATTTACAGAAAGAACAGAGGATCATGTGTATCATCCCAACAGGAGCTGCGAGTGATA 548
Qy 481 GCTGGACAAGGACAATTTGCCCTGGAAGTGTGTAACAGGTTCTTTGGTGGATGCACTG 540
Db 549 GCTGGACAAGGACAATTTGCCCTGGAAGTGTGTAACAGGTTCTTTGGTGGATGCACTG 608
Qy 541 GTGCTACTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTCTG 600
Db 609 GTGCTACTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTCTG 668
Qy 601 AAACCTAGTGTGAAGGTATATGCTGTAACCTCAATCAATCAGATGACTGCTACCAGTCC 660
Db 669 AAACCTAGTGTGAAGGTATATGCTGTAACCTCAATCAATCAGATGACTGCTACCAGTCC 728
Qy 661 AAGCTGAAGGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGTTGTC 720
Db 729 AAGCTGAAGGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGTTGTC 788
Qy 721 AAATCCAGCATTTGCTTGAACCTGCGCTTATATCAGGAGACCTTGTGGAATGATATCTTC 780
Db 789 AAATCCAGCATTTGCTTGAACCTGCGCTTATATCAGGAGACCTTGTGGAATGATATCTTC 848
Qy 781 ACTGTACACAGATGAATTAAGTGTGCAACCTGCTGTGGGAGAGGATGAACCTA 840
Db 849 ACTGTACACAGATGAATTAAGTGTGCAACCTGCTGTGGGAGAGGATGAACCTA 908
Qy 841 CTCATTGAACCTACAGCTGCTGTGGAGTGGCTGTGCTGCTCAACATTTTCAAACT 900
Db 909 CTCATTGAACCTACAGCTGCTGTGGAGTGGCTGTGCTGCTCAACATTTTCAAACT 968
Qy 901 GTTTCCTCCAGAACTTAAGATTTGCTGCTAGTGGTGAATGATGACTTAACC 960
Db 969 GTTTCCTCCAGAACTTAAGATTTGCTGCTAGTGGTGAATGATGACTTAACC 1028
Qy 961 TCCTCCATAACTTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTCTGTT 1020
Db 1029 TCCTCCATAACTTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTCTGTT 1088
Qy 1021 TAA 1023
Db 1089 TAA 1091

RESULT 4
AK023169
LOCUS
DEFINITION
Homo sapiens CDNA FLJ13107 fis, clone NT2RP3002501, weakly similar
to THREONINE DEHYDRATASE CATABOLIC (EC 4.2.1.16).
ACCESSION
AK023169
VERSION
AK023169.1 GI:10434971
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens
clone lib: NT2RP3 clone: NT2RP3002501.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,
Wagatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,
Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2477)
TITLE
AGAGTTTACAGAGAAACAGAGGCATCATGTTACATCCCAACAGGAGCTGCGAGTGATA 480
JOURNAL
Db 489 AGAGTTTACAGAGAAACAGAGGCATCATGTTACATCCCAACAGGAGCTGCGAGTGATA 548
REFERENCE
AUTHORS
Isogai, T. and Otsuki, T.
```

```
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction; 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
1..2477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3002501"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP3"
/notes="cloning vector: pME18SFL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
CDS
69..1091
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB14442.1"
/db_xref="GI:10434972"
/translation="MCAQYICISFADVEKAHINIRDSIHLTPVLVTSSILNLTGRNLF
KSELFQRTSFKIRGALNVSLPDALEKRKPAVYVTHSSNGHQAITYAAKLEGP
YIVYVQTPADCKKLAIGAYGASTVYCEPSEDSKRVAKRVEETEGIMVAPSP
AGCCTALEVLNQPLVDALVVPVGGMLAGIATVKALKPVKVVAAKPSNADDCY
QSKLKAKMPLNLYPPTIADQVKSISGLNTPIRLVDIDFTVTEDEIKCATQLV
RMKLLIETPAGVGAVALSOHFQTVSPVKNICIVLSSGNVDLTSSITWVQAE
RPAAS
YQSVSV"
BASE COUNT 738 a 535 c 524 g 680 t
ORIGIN
Query Match 100.0%; Score 1023; DB 9; Length 2477;
Best Local Similarity 100.0%; Pred. No. 6.3e-266;
Matches 1023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGTGTGCTCAGTATTGTCATCTCTTTCGTGATGTTGAAAAGCTCATATCAACATTCGA 60
Db 69 ATGTGTGCTCAGTATTGTCATCTCTTTCGTGATGTTGAAAAGCTCATATCAACATTCGA 128
Qy 61 GATTCTATCCACCTCACACAGTCTTCCAGAAAACAGGATCTTTTAAGATTCGTGGTGT 120
Db 129 GATTCTATCCACCTCACACAGTCTTCCAGAAAACAGGATCTTTTAAGATTCGTGGTGT 188
Qy 121 AATCTTTTCTTCAATGTGAACCTCTTCCAGAAAACAGGATCTTTTAAGATTCGTGGTGT 180
Db 189 AATCTTTTCTTCAATGTGAACCTCTTCCAGAAAACAGGATCTTTTAAGATTCGTGGTGT 248
Qy 181 CTCATATCCGCTCAGAAGCTTGGTTCCTGATGCTTTTGAAGAGAGCGGAAAGCTGTGTT 240
Db 249 CTCATATCCGCTCAGAAGCTTGGTTCCTGATGCTTTTGAAGAGAGCGGAAAGCTGTGTT 308
Qy 241 ACTCACAGCAGTGGAAACCATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGAAAT 300
Db 309 ACTCACAGCAGTGGAAACCATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGAAAT 368
Qy 301 CTGCTTTATATTGTTGGTGGCCCAACAGAGCTTCCAGAGCTGTAAAAAAGCTTGAATACAGCC 360
Db 369 CTGCTTTATATTGTTGGTGGCCCAACAGAGCTTCCAGAGCTGTAAAAAAGCTTGAATACAGCC 428
Qy 361 TAGCGAGCGTCAATTTGATGTAACCTAGTGTAGAGTCCAGAGAAAATGTTGCAAAA 420
Db 429 TAGCGAGCGTCAATTTGATGTAACCTAGTGTAGAGTCCAGAGAAAATGTTGCAAAA 488
Qy 421 AGAGTTTACAGAGAAACAGAGGCATCATGTTACATCCCAACAGGAGCTGCGAGTGATA 480
Db 489 AGAGTTTACAGAGAAACAGAGGCATCATGTTACATCCCAACAGGAGCTGCGAGTGATA 548
```

Qy	481	GCTGGACAAGGGACAATTGCCCTGGAAGTCTGCTGAACACGAGTTCCTTTGGTGGATGCACCTG	540
Db	549	GCTGGACAAGGGACAATTGCCCTGGAAGTCTGCTGAACACGAGTTCCTTTGGTGGATGCACCTG	608
Qy	541	GTGGTACCTGTAGGTGGAGGAGGAATGCTTGGCTTGGAAATAGCAATTAAGTTAAGGCTCTG	600
Db	609	GTGGTACCTGTAGGTGGAGGAGGAATGCTTGGCTTGGAAATAGCAATTAAGTTAAGGCTCTG	668
Qy	601	AAACCTAGTGTGAAGGTATATGCTGTGTAACCCCTCAAATGCAGATGACTGCTACCAGTCC	660
Db	669	AAACCTAGTGTGAAGGTATATGCTGTGTAACCCCTCAAATGCAGATGACTGCTACCAGTCC	728
Qy	661	AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGT	720
Db	729	AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGT	788
Qy	721	AAATCCAGCAATTCGCTTGAACACCTGGCCTATTATCAGGAGCCTTGTGGATGATATCTTC	780
Db	789	AAATCCAGCAATTCGCTTGAACACCTGGCCTATTATCAGGAGCCTTGTGGATGATATCTTC	848
Qy	781	ACTGTCCACAGGATGAAATTAAGTGTCAACCCAGCTGGTGTGGAGAGGATGAAACTA	840
Db	849	ACTGTCCACAGGATGAAATTAAGTGTCAACCCAGCTGGTGTGGAGAGGATGAAACTA	908
Qy	841	CTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT	900
Db	909	CTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT	968
Qy	901	GTATCCCCAGAACTAAGAACATTTGATTTGCTCTCAGTGGTGGAAATGTAGACTTAACC	960
Db	969	GTATCCCCAGAACTAAGAACATTTGATTTGCTCTCAGTGGTGGAAATGTAGACTTAACC	1028
Qy	961	TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTTCTGTT	1020
Db	1029	TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTTCTGTT	1088
Qy	1021	TAA 1023	
Db	1089	TAA 1091	
RESULT 5			
AX259575			
LOCUS	AX259575	1336 bp	DNA linear PAT 26-OCT-2001
DEFINITION	Sequence 1 from Patent WO0173077.		
ACCESSION	AX259575		
VERSION	AX259575.1 GI:16508702		
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1		
AUTHORS	Ramakrishnan,S.		
TITLE	Regulation of human serine racemase enzyme		
JOURNAL	Patent: WO 0173077-A 1 04-OCT-2001;		
FEATURES	Bayer Aktiengesellschaft (DE)		
source	Location/Qualifiers		
	1..1336		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	386 a	268 c	319 g 362 t 1 others
ORIGIN			
Query Match 99.8%; Score 1021.4; DB 6; Length 1336;			
Best Local Similarity 99.9%; Pred. No. 1.8e-265;			
Matches 1022; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Qy	1	ATGTGTGCTCAGTATTGCTATCTCTTCTGCTGATGTTGAAAAGGCTCATATCAACATTCGA	60
Db	52	ATGTGTGCTCAGTATTGCTATCTCTTCTGCTGATGTTGAAAAGGCTCATATCAACATTCGA	111
Qy	61	GATTCTATCCACCTCACACCAAGTCTTAACAAGCTCCATTTTGAATCAACTTAACAGGGCGC	120

Db	112	GATTCTATCCACCTCACACCAAGTCTTAACAAGCTCCATTTTGAATCAACTAACAGGGCGC	171
QY	121	AATCTTTTCTTCAATGTGAAGTCTTCCAGAAAACAGGATCTTTTAAGATTCTGGTGCT	180
Db	172	AATCTTTTCTTCAATGTGAAGTCTTCCAGAAAACAGGATCTTTTAAGATTCTGGTGCT	231
QY	181	CTCAATGCCGTGCAAGCTTGGTCTCTGATGCTTTAGAAAAGGAAAGCCGAAAGCTGTCTGT	240
Db	232	CTCAATGCCGTGCAAGCTTGGTCTCTGATGCTTTAGAAAAGGAAAGCCGAAAGCTGTCTGT	291
QY	241	ACTCACAGCAGTGAAGAACCATGGCCAGGCTCTACCTATATGCTGCCAAATTTGGAAGGAAT	300
Db	292	ACTCACAGCAGTGAAGAACCATGGCCAGGCTCTACCTATATGCTGCCAAATTTGGAAGGAAT	351
QY	301	CCTGCTTATATTTGGTGGTCCCCCAGACACTCCAGACGTAAAAAACTTGCAATACAAGCC	360
Db	352	CCTGCTTATATTTGGTGGTCCCCCAGACACTCCAGACGTAAAAAACTTGCAATACAAGCC	411
QY	361	TACGAGGCGTCAATTTGATGATGAGTCCAGAGAAAATGTTGCAAAA	420
Db	412	TACGAGGCGTCAATTTGATGATGAGTCCAGAGAAAATGTTGCAAAA	471
QY	421	AGAGTTACAGAAAACAGAAAGGATCATGTGTATATCCCAACAGGAGCCCTGCAGTGATA	480
Db	472	AGAGTTACAGAAAACAGAAAGGATCATGTGTATATCCCAACAGGAGCCCTGCAGTGATA	531
QY	481	GCTGGACAAGGGCAATTTGCCCTGGAAGTGTGAACAGGTTCTTTGGTGGATGCACTG	540
Db	532	GCTGGACAAGGGCAATTTGCCCTGGAAGTGTGAACAGGTTCTTTGGTGGATGCACTG	591
QY	541	GTGCTACCTGTAGTGGAGGAGGAATGCTTCTGGAATAGCAATTTACAGTTAAGGCTCTG	600
Db	592	GTGCTACCTGTAGTGGAGGAGGAATGCTTCTGGAATAGCAATTTACAGTTAAGGCTCTG	651
QY	601	AAACCTAGTGTGAAGGTATATGCTGTGAACCCCTCAAAATGCAGATGACTGTACCAGTCC	660
Db	652	AAACCTAGTGTGAAGGTATATGCTGTGAACCCCTCAAAATGCAGATGACTGTACCAGTCC	711
QY	661	AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCTC	720
Db	712	AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCTC	771
QY	721	AAATCCAGCATTGGCTTGAACACCTGGCTTATATCAGGAGCCTTTGTGGATGATATCTTC	780
Db	772	AAATCCAGCATTGGCTTGAACACCTGGCTTATATCAGGAGCCTTTGTGGATGATATCTTC	831
QY	781	ACTGTACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGGAGAGGATGAAACTA	840
Db	832	ACTGTACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGGAGAGGATGAAACTA	891
QY	841	CTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGCTGTGCTGTCTCAACATTTTCAAACT	900
Db	892	CTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGCTGTGCTGTCTCAACATTTTCAAACT	951
QY	901	GTTCCTCCAGAAAGTAAGAACATTTGATTTGCTGCTCAGTGTGGGAAATGTAGACTTAACC	960
Db	952	GTTCCTCCAGAAAGTAAGAACATTTGATTTGCTGCTCAGTGTGGGAAATGTAGACTTAACC	1011
QY	961	TCCTCCATAACTTGGGTGAAGCAGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGTT	1020
Db	1012	TCCTCCATAACTTGGGTGAAGCAGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGTT	1071
QY	1021	TAA 1023	
Db	1072	TAA 1074	
RESULT 6			
LOCUS	AX224420	1020 bp	DNA linear PAT 11-SEP-2001
DEFINITION	Sequence 3 from Patent WO0160987.		
ACCESSION	AX224420		

```
VERSION      AX224420.1  GI:15554662
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE        1 (bases 1 to 1020)
JOURNAL      A human pyridoxal-phosphate dependent enzyme family member and uses
FEATURES     Patent: WO 0160987-A 3 23-AUG-2001;
              Millennium Pharmaceuticals, Inc. (US)
              Location/Qualifiers
              1..1020
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
BASE COUNT   288 a 220 c 243 g 269 t
ORIGIN
Query Match      99.7%; Score 1020; DB 6; Length 1020;
Best Local Similarity 100.0%; Pred. No. 4.3e-265;
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ATGTCGTCAGTATGTCATCTCTTCTGTCGATGTTGAAAGAGCTCATATCAACATTTCGA 60
DB      1 ATGTCGTCAGTATGTCATCTCTTCTGTCGATGTTGAAAGAGCTCATATCAACATTTCGA 60
QY      61 GATTCATATCCACCTCACACAGTGTCTAACAAGCTCCATTGTAATCAACTAACAGGCGC 120
DB      61 GATTCATATCCACCTCACACAGTGTCTAACAAGCTCCATTGTAATCAACTAACAGGCGC 120
QY      121 AATCTTTTCTTCAATGTGAATCTCTCCAGAAACAGGATCTTTTAAAGATTCGTGGTCT 180
DB      121 AATCTTTTCTTCAATGTGAATCTCTCCAGAAACAGGATCTTTTAAAGATTCGTGGTCT 180
QY      181 CTCATATGCGTCAGAGCTGTTGTTCTCTGATGCTTTAGAAAGAGCCGAAAGCTTGTGT 240
DB      181 CTCATATGCGTCAGAGCTGTTGTTCTCTGATGCTTTAGAAAGAGCCGAAAGCTTGTGT 240
QY      241 ACTCACAGCTGGAACCATGCGCCAGCTCTACCTATGCTGCCAAATTTGAAGGAATT 300
DB      241 ACTCACAGCTGGAACCATGCGCCAGCTCTACCTATGCTGCCAAATTTGAAGGAATT 300
QY      301 CTGCTTATATGTTGGTCCCCAGACAGCTCCAGACTGTAAACAACTTGCATACAGCC 360
DB      301 CTGCTTATATGTTGGTCCCCAGACAGCTCCAGACTGTAAACAACTTGCATACAGCC 360
QY      361 TACGAGCGTCATTTGTATCTGTAACCTAGTAGTATGATCCAGAGAAATGTTGCAAAA 420
DB      361 TACGAGCGTCATTTGTATCTGTAACCTAGTAGTATGATCCAGAGAAATGTTGCAAAA 420
QY      421 AGAGTTACAGAGAAACAGAGGATCATGTTGATATCCCAACAGGAGCCCTGCAGTGATA 480
DB      421 AGAGTTACAGAGAAACAGAGGATCATGTTGATATCCCAACAGGAGCCCTGCAGTGATA 480
QY      481 GCTGACAAAGGACAAATTCCTGGAAGTGTGTAACACAGGTTCTCTTGGTGGATGCACTG 540
DB      481 GCTGACAAAGGACAAATTCCTGGAAGTGTGTAACACAGGTTCTCTTGGTGGATGCACTG 540
QY      541 GTGGTACCTGTAGGTGGAGGAGGATGCTTCTGGAATAGCAATACAGTTAAGGCTGTG 600
DB      541 GTGGTACCTGTAGGTGGAGGAGGATGCTTCTGGAATAGCAATACAGTTAAGGCTGTG 600
QY      601 AAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAATATGCAGATGACTGTACCAAGTCC 660
DB      601 AAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAATATGCAGATGACTGTACCAAGTCC 660
QY      661 AAGCTGAAGGGAAGAACTGATGCCCAATCTTTATCTCCAGAAACCATACAGATGGTGTG 720
DB      661 AAGCTGAAGGGAAGAACTGATGCCCAATCTTTATCTCCAGAAACCATACAGATGGTGTG 720
QY      721 AAATCCAGCATTTGGCTTCAACACCTGGCTTATATCAGGAGCCTTGTGGATGATATCTTC 780
DB      721 AAATCCAGCATTTGGCTTGAACACCTGGCTTATATCAGGAGCCTTGTGGATGATATCTTC 780
QY      781 ACTGTCACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGGATGAAACTA 840
DB      781 ACTGTCACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGGATGAAACTA 840
QY      841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTCTCAACATTTTCAAACT 900
DB      841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTCTCAACATTTTCAAACT 900
QY      901 GTTTCCCCAGAGTAAGAACATTTGTATTTGTCTCAGTGGTGGAAATGTAGACTTAACC 960
DB      901 GTTTCCCCAGAGTAAGAACATTTGTATTTGTCTCAGTGGTGGAAATGTAGACTTAACC 960
QY      961 TCCTCATAAATTTGGTGAAGCAGCTGAAAGCCAGCTTCTTATCAGTCTGTTCTGTT 1020
DB      961 TCCTCATAAATTTGGTGAAGCAGCTGAAAGCCAGCTTCTTATCAGTCTGTTCTGTT 1020
RESULT 7
LOCUS       HSM805450          2264 bp      mRNA      linear      PRI 12-JUL-2002
DEFINITION Homo sapiens mRNA; cDNA DKFZp762A2415 (from clone DKFZp762A2415).
ACCESSION   AL834378
VERSION     AL834378.1      GI:21740048
KEYWORDS    human.
SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE   1 (bases 1 to 2264)
AUTHORS     Ansoorge,W., Wirkner,U., Mewes,H.W., Weill,B. and Wiemann,S.
TITLE       Direct Submission
JOURNAL     Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp762A2415) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
             1..2264
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /map="22.9 cR from top of Chr17 linkage group"
                /clone="DKFZp762A2415"
                /tissue_type="melanoma (MeWo cell line)"
                /clone_lib="762 (synonym: hmcl2). Vector pSport1; host
                DH10B; sites NotI + SalI"
                /dev_stage="adult"
             1..2264
                /gene="DKFZp762A2415"
                /c1..867
                /note="similarity to serine racemase (Mus musculus)"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAD39041.1"
                /db_xref="GI:21740049"
                /translation="GSPKRGALNARSLVDPALERPKAVVTHSSGNHGOALTVAAK
                LEGIPAVIVPQATPCKKLAIOAYGASIVYCEPDESRENNVAKRVTEETEGIMVHPN
                QEPVIAQGIHTALEVLNOVELDALVVPVGGGMLAGIATVTKALRFSVKVYAAPS
                NADCYCKLKLKLPNLYPPETADGVKSSIGLNTWPIIRDLVDIDFTVTEDEIKCA
                TOLYMERKLLIEPTAGVGVAAVLQHFQTVSPVKNICIVISGNVDLTSSITWVKQ
                AERPASQSVSV"
                2226..2231
                /gene="DKFZp762A2415"
                2253
                /gene="DKFZp762A2415"
polyA_signal
polyA_site
```



Db 486 AGAATTATGCAAGAAACAGAGGATCTTGGTCCATCCCAACAGGAGCCTGCAGTGATA 545  
 QY 481 GCTGGACAAGGACAAATGCTGGAAGTCTGTAACCAAGGTTCTTTGGTGGATGCACTG 540  
 Db 546 GCTGGACAAGGACAAATGCTGGAAGTCTGTAACCAAGGTTCTTTGGTGGATGCACTG 605  
 QY 541 GTGGTACCTGTAGTGGAGGAGGAATGCTTGGTAATAGCAATACAGTAAAGGCTCTG 600  
 Db 606 GTGGTACCTGTAGGAGGAGGAGGAATGCTTGGTAATAGCAATACAGTAAAGGCTCTG 665  
 QY 601 AAACCTAGTGTAGAGGTATATGCTGTGAACCTCAATGCAAGATGCTTACCAGTCC 660  
 Db 666 AAACCTAGTGTAGAGGTATATGCTGTGAACCTCAATGCAAGATGCTTACCAGTCT 725  
 QY 661 AAGCTGAAGGGGAACCTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCT 720  
 Db 726 AAACCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 785  
 QY 721 AAATCCAGCATTTGGCTTTGAACACCTGGCCCTATTATCAGGAGCCTTGTGGATGATCTTC 780  
 Db 786 AAATCCAGCATTTGGCTTTGAACACCTGGCCCTATTATCAGGAGCCTTGTGGATGATCTTC 845  
 QY 781 ACTGTCCACAGAGATGAATTAAGTGTGAACCCAGCTGTGTGGGAGAGGATGAACACTA 840  
 Db 846 ACTGTCCACAGAGATGAATTAAGTGTGAACCCAGCTGTGTGGGAGAGGATGAACACTA 905  
 QY 841 CTCAATGAACCTACAGCTGTGTGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900  
 Db 906 CTCAATGAACCTACAGCTGTGTGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 965  
 QY 901 GTTTCCTCCAGAGTGAAGAAACATTTGTTATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 960  
 Db 966 GTTTCCTCCAGAGTGAAGAAACATTTGTTATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1023  
 QY 961 TCCTCCATTAATCTGGGTGAAGAGGCTGAAGGCCAGCTTCTTATCAGTCTGTCTGTCTGTT 1020  
 Db 1024 -CTCCTCTGAACCTGGGTGGGAGGCTGAAGGCCAGCTTCTTACCAGAGCGTTTCTGTT 1082  
 QY 1021 TAA 1023  
 Db 1083 TAA 1085  
 BC011164 1411 bp mRNA linear ROD 07-AUG-2002  
 Mus musculus, Similar to serine racemase, clone MGC:18670  
 IMAGE:4195695, mRNA, complete cds.  
 BC011164  
 MGC.  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1411)  
 Strausberg, R.  
 Direct Submission  
 Submitted (25-JUL-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,  
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 24 Row: m Column: 6  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 7305520.

FEATURES  
 source  
 location/Qualifiers  
 1..1411  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="FVB/N"  
 /clone="MGC:18670 IMAGE:4195695"  
 /tissue\_type="Liver, normal. 5 month old male mouse."  
 /clone\_lib="NCI\_CGAP\_L19"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 288..1307  
 /codon\_start=1  
 /product="Similar to serine racemase"  
 /protein\_id="AAH11164.1"  
 /db\_xref="GI:15029881"  
 /translation="MCAQYICISFADYVEKAHINIQDSIHLTPVLTSSILNOIAGRNLF  
 KCELFQKTSFKIRGALNAILPDTPEKPKAVVTHSHGNHQAITYAAKLEGI  
 YIVVPTANCKKLAIQNGASIVYCDPSDESEKVTQRTQIMQETGILVHPNEPAV  
 QGKLTALVNLQVPLVDALVPVGGGMMVAGIATIKALKPSVKVYAAEPNADCY  
 ASKLKELTPNLHPPTIADLVKSSIGLNTWPIIRDLVDVFTVDEIKYATQLVYMG  
 RKLLIETAGVALAAVLQHFQTVSPVKNVCIVLSGGNVLDLSLNVVQAEAPY  
 QTQVSV"  
 BASE COUNT 411 a 335 c 340 g 325 t  
 ORIGIN

Query Match 79.8%; Score 816.6; DB 10; Length 1411;  
 Best Local Similarity 88.1%; Pred. No. 4.9e-210;  
 Matches 901; Conservative 0; Mismatches 119; Indels 3; Gaps 1;  
 QY 1 ATGTGTGTCAGTATTGTCATCTCTTGTGTCATGTTGAAAAGCTCATATCAACATCGA 60  
 Db 288 ATGTGTGTCAGTATTGTCATCTCTTGTGTCATGTTGAAAAGCTCATATCAACATCGA 347  
 QY 61 GATTCTATCCACCTCACACAGCTGCTAAAGCTCCATTTTGAATCAACTTAACAGGCGC 120  
 Db 348 GACTCTATCCACCTCACACAGCTGCTAAAGCTCCATTTTGAATCAACTTAACAGGCGC 407  
 QY 121 AATCTTTTCTTCAAAATGTGAACCTTTCAGAAACAGGATCTTTTAAAGTTCGTGGTCT 180  
 Db 408 AATCTTTTCTTCAAAATGTGAACCTTTCAGAAACAGGATCTTTTAAAGTTCGTGGTCT 467  
 QY 181 CTCAATGCCGTCAGAAAGCTTGGTCTCTGATGCTTTAGAAGGAGCGGAAGCTGTGTT 240  
 Db 468 CTTAATGCCATCAGAGGCTTAATCTCTGACACCCAGAGAGAGGCGGAGTGT 527  
 QY 241 ACTCACAGGAGTGGAAACCATGGCCAGGCTCTCACCTATCTGCCAAATTTGAAGAAAT 300  
 Db 528 ACTCACAGGAGTGGAAACCATGGCCAGGCTCTCACCTATCTGCTAACTGGAGGAAT 587  
 QY 301 CCGTCTTATTTGTTGTGCCCCAGACAGCTGCCAGACTGTAAACAACTTGCATACAGCC 360  
 Db 588 CCGTCTTATTTGTTGTGCCCCAGACAGCTGCCAGACTGTAAACAACTTGCATACAGCC 647  
 QY 361 TAGGGAGCTCAATTTGATCTGTAACCTAGTAGTGTGAGTCCAGAGAAATTTGTCAAA 420  
 Db 648 TAGGGAGCTCAATTTGATCTGTAACCTAGTAGTGTGAGTCCAGAGAAATTTGTCAAA 707  
 QY 421 AGAGTTTACAGAAAGAGAGGATCATGTTGATCCCAACAGGAGGCTGCAGTGTATA 480  
 Db 708 AGAATTATCCAGAAACAGAGGATCTTTGGTCCATCCCAACAGGAGGCTGCAGTGTATA 767  
 QY 481 GCTGGACAAGGAGCAATTCCTCTGGAAGTGTGTAACAGGATCTTTTGTGGTGGATGCTG 540



XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro  
 DNA Res. 7 (1), 65-73 (2000)  
 20181126  
 2 (bases 1 to 4107)  
 Direct Submission  
 Ohara, O., Nagase, T. and Kikuno, R.  
 Submitted (31-JAN-2000) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan (E-mail:cdna@kzusa.or.jp, URL: http://www.kazusa.or.jp/huge/, Tel: +81-438-52-3913, Fax: +81-438-52-3914)

## FEATURES

Location/Qualifiers

Source

1..4107  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="hk09639"  
 /tissue\_type="brain"  
 /clone\_lib="pBluescriptII SK plus"  
 1..4107  
 /gene="KIAA1401"  
 /gene="KIAA1401"  
 /note="Start codon is not identified."  
 /codon\_start=1  
 /product="KIAA1401 protein"  
 /protein\_id="BAA92639.1"  
 /db\_xref="GI:7243183"  
 /translation="HGALTSCKWRPSREGAGSPAPWRARPLGGDKRLKRLSWPSSADSV  
 HARGMAARPGPLKQNKAKHGGRGSAORGDGKRLALKLTSKVKRKLRSVDQ  
 RRASOLRKKKEAVLAELKQLGKDGKPPHQLVYVPLHSRLSPLEAMQLQDRTGV  
 HNLGELTQNPMLCPLKRLKRFETSAKPPGDLVVLDMKAVDTILFLDPLEGWDST  
 GDYCLSLFAOGLTYTLAVOGISGLPKKQIDTRKLSKAVKRRFPDRLKLLDQTLQ  
 EAGMLQLANLQKOOHLAFRRAYLFAHAFVDFPSENNLVGTLKISGVVGGTLN  
 NLLHIVGYDQFQKIDAPDPPNPGIKPKQDPDAMEICATDAVDMDMERGLAV  
 LMKADPQBSLQFVLPDMEGPTWPTTEELSEAKDFLKESSKVKYKPKGTSSSQ  
 AEWLDDGSGSGGDEYVDMDHEFDEFEESODESESEEEYETMTIGESVHDDLY  
 DKVDEAEAKMLKYLKQRLKLEENFDEVDTPRDAARIRFKYRGLKSPSPWDFQ  
 ENLPDFAIRLQFNFTNRKRSFKEVEEVEGEAGVYVTLHVSFVPSVVECFRQ  
 GTPLIAFLSLPHBQKMSVLNVRDRDNPTEPPASVLIFKOKSNMHSLIAHGLMSVDPDR  
 AADKKLQRELTADMLAVATVYAPITPPASVLIFKOKSNMHSLIAHGLMSVDPDR  
 MVTKRVYLSHPHFPIFTKMAVRYMREDDVLFKPELRTKMGWGHILKEPLTHG  
 HMKCSFGDKLSODTFLNMLYKRVFEPKWTYDYPYEPVPLKSEISSTVPPQGMG"

gene

CDS

1094 a 992 c 1131 g 890 t

BASE COUNT

ORIGIN

Query Match 21.68; Score 221; DB 9; Length 4107;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-49;  
 Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 803 AGTGTGCAACCCAGCTGGTGGGAGAGGATGAACCTACTTGAACCTACAGCTGGT 862  
 Db 4037 AGTGTGCAACCCAGCTGGTGGGAGAGGATGAACCTACTTGAACCTACAGCTGGT 922  
 QY 863 TTGGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTTCCCGAGAAGTAAAGACA 922  
 Db 3977 TTGGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTTCCCGAGAAGTAAAGACA 3918  
 QY 923 TTGTGATTGCTCAGTGGTGGAAATGTAGACTTAACCTCCTCCATAAATTTGGGTGAAGC 982  
 Db 3917 TTGTGATTGCTCAGTGGTGGAAATGTAGACTTAACCTCCTCCATAAATTTGGGTGAAGC 3858  
 QY 983 AGCTGAAAGCCAGCTCTTATCAGTCTCTTCTGTTTAA 1023  
 Db 3857 AGCTGAAAGCCAGCTCTTATCAGTCTCTTCTGTTTAA 3817

RESULT 12  
 AK026565/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

AK026565  
 Homo sapiens cDNA: FLJ22912 fis, clone KAT06132.  
 AK026565.1 GI:10439447

AK026565  
 Homo sapiens  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

oligo capping: fis (full insert sequence).  
 Homo sapiens signet-ring cell carcinoma cell line:KATO III cDNA to mRNA, clone lib:KAT clone:KAT06132.  
 Homo sapiens  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 1 (sites)  
 Watanabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H.,  
 Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,  
 Nakamura, Y., Isogai, T. and Sugano, S.  
 NEDO human cDNA sequencing project  
 2 (bases 1 to 4140)  
 Submitter  
 Shibahara, T., Tanaka, T. and Nakamura, Y.  
 Direct Submission  
 Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
 University of Tokyo, Laboratory of Genome Structure Analysis, Human  
 Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
 Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
 Fax:81-3-5449-5416)  
 NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan; cDNA full insert  
 sequencing; Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing; Department of  
 Virology and Human Genome Center, Institute of Medical Science,  
 University of Tokyo (partly supported by Science and Technology  
 Agency).

Location/Qualifiers  
 1..4140  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="KAT06132"  
 /cell\_line="KATO III"  
 /cell\_type="signet-ring cell carcinoma"  
 /clone\_lib="KAT"  
 /note="cloning vector pME18SFL3"  
 1195 a 901 c 1050 g 994 t

BASE COUNT

ORIGIN

Query Match 21.68; Score 221; DB 9; Length 4140;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-49;  
 Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 803 AGTGTGCAACCCAGCTGGTGGGAGAGGATGAACCTACTTGAACCTACAGCTGGT 862  
 Db 2988 AGTGTGCAACCCAGCTGGTGGGAGAGGATGAACCTACTTGAACCTACAGCTGGT 2929  
 QY 863 TTGGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTTCCCGAGAAGTAAAGACA 922  
 Db 2928 TTGGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTTCCCGAGAAGTAAAGACA 2869  
 QY 923 TTGTGATTGCTCAGTGGTGGAAATGTAGACTTAACCTCCTCCATAAATTTGGGTGAAGC 982  
 Db 2868 TTGTGATTGCTCAGTGGTGGAAATGTAGACTTAACCTCCTCCATAAATTTGGGTGAAGC 2809  
 QY 983 AGCTGAAAGCCAGCTCTTATCAGTCTCTTCTGTTTAA 1023  
 Db 2808 AGCTGAAAGCCAGCTCTTATCAGTCTCTTCTGTTTAA 2768

RESULT 13  
 HSB17A99/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

HSB17A99  
 Homo sapiens chromosome 17, sequence from PAC RPCI-5 1037N22 map  
 AL450226 AL137039  
 AL450226.1 GI:11228434  
 HTG.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE
AUTHORS Ramser,J., Langer,I., Klein,M., Arndt,S., Lehrack,S., Junker,E.,
Schuelzchen,S., Noitz,M., Thompson,C., Hoff,C., Poustka,A.,
Reinhardt,R. and Lehrach,H.
JOURNAL
REFERENCE
AUTHORS MOLGENR.
TITLE Direct Submission
JOURNAL
COMMENT Submitted (20-JAN-2000) MOLGENR, Abt. Lehrach, Max Planck Institut
Fuhr Molekulare Genetik, Innestrasse 73, Berlin, 14195 Germany
contig 01 1. .123454.
FEATURES
source
1. .123454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q13.3"
/clone="PAC RPCI-5 1037N22"
/clone_lib="RPCI1,3-5, Roswell Park Cancer Institute,
creator Pieter de Jong, P.Ioannou"
/notes="region between markers D17S695 - D17S654"
1
misc_feature
/notes="T7_end:PAC RPCI-5 1037N22"
123454
misc_feature
/notes="SP6_end:PAC RPCI-5 1037N22"
31233 a 26833 c 28651 g 36737 t
BASE COUNT
ORIGIN
Query Match 21.6%; Score 221; DB 9; Length 123454;
Best Local Similarity 100.0%; Pred. No. 7e-49; Indels 0; Gaps 0;
Matches 221; Conservative 0; Mismatches 0;

QY 803 AGTGTCAACCCAGCTGGTGGGAGGATGAACCTACTTGAACCTACAGCTGGTG 862
|||||
Db 15555 AGTGTCAACCCAGCTGGTGGGAGGATGAACCTACTTGAACCTACAGCTGGTG 15496
|||||

QY 863 TTGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTCCCGAGAGTAAGAACA 922
|||||
Db 15495 TTGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTCCCGAGAGTAAGAACA 15436
|||||

QY 923 TTGTATTGTGCTCACTGGTGGAAATGTAGACTTAACCTCCCTCATTAACCTGGTGAAGC 982
|||||
Db 15435 TTGTATTGTGCTCACTGGTGGAAATGTAGACTTAACCTCCCTCATTAACCTGGTGAAGC 15376
|||||

QY 983 AGGCTGAAGGCGAGCTTCTTATCAGTCTGTTTCTGTTAA 1023
|||||
Db 15375 AGGCTGAAGGCGAGCTTCTTATCAGTCTGTTTCTGTTAA 15335
|||||

RESULT 14
AC021705/c
LOCUS 221909 bp DNA linear HTG.23-AUG-2002
DEFINITION Homo sapiens chromosome 17 clone RP11-380H7 map 17, *** SEQUENCING
IN PROGRESS ***, 5 unordered pieces.
AC021705
ACCESSION AC021705.19 GI:22450645
VERSION HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 221909)
Birken,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone RP11-380H7
Unpublished
2 (bases 1 to 221909)
Birken,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bida,F.,
Boquslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepl,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE Direct Submission
JOURNAL
COMMENT Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 221909)
Birken,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepl,Y., Collymore,A., S.,
Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Minova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 23, 2002 this sequence version replaced gi:22296935.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5879
Center clone name: 380_H_7
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 49592: contig of 49592 bp in length
* 49593 49692: gap of 100 bp
* 49693 65686: contig of 15994 bp in length
* 65687 65786: gap of 100 bp
* 65787 69248: contig of 3462 bp in length
* 69249 69348: gap of 100 bp
* 69349 92592: contig of 23244 bp in length
* 92593 92692: gap of 100 bp
* 92693 221909: contig of 129217 bp in length.
*
* Location/Qualifiers
* 1. .221909
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="17"
* /map="17"
* /clone="RP11-380H7"
* /clone_lib="RPCI-11 Human Male BAC"

```

BASE COUNT 53146 a 52419 c 54384 g 61545 t 415 others  
ORIGIN

Query Match 21.6%; Score 221; DB 2; Length 221909;  
Best Local Similarity 100.0%; Pred. No. 6.8e-49;  
Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 803 AGTGTGCAACCCAGCTGGTGGGAGAGGATGAACCTACTCATTTGAACCTACAGCTGGTG 862  
Db 74677 AGTGTGCAACCCAGCTGGTGGGAGAGGATGAACCTACTCATTTGAACCTACAGCTGGTG 74618  
Qy 863 TTGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTTCCCCAGAAAGTAAAGAACA 922  
Db 74617 TTGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTTCCCCAGAAAGTAAAGAACA 74558  
Qy 923 TTTCTATTGCTCAGTGGTGGGAATGTAGACTTAACCTCTCCATAACTTGGGTGAAGC 982  
Db 74557 TTTCTATTGCTCAGTGGTGGGAATGTAGACTTAACCTCTCCATAACTTGGGTGAAGC 74498  
Qy 983 AGGCTGAAGGCCAGCTCTTATCAGTCTGTTCTGTTTAA 1023  
Db 74497 AGGCTGAAGGCCAGCTCTTATCAGTCTGTTCTGTTTAA 74457

RESULT 15  
AL604066/c  
LOCUS  
DEFINITION Mouse DNA sequence from clone Rp23-174M12 on chromosome 11,  
complete sequence.  
ACCESSION AL604066  
VERSION AL604066.9 GI:19571943  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 184151)  
Bray-Allen, S.  
Direct Submission  
Submitted (14-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk  
On Mar 21, 2002 this sequence version replaced gi:17902958.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em., EMBL; Sw.,  
SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
from the RPI-23 Mouse PAC Library  
constructed by the group of Pieter de Jong.  
For further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6

This sequence is the entire insert of clone Rp23-174M12 The true  
right end of clone Rp23-194P5 is at 82810 in this sequence.

FEATURES  
source  
Location/Qualifiers  
1..184151  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/clone="Rp23-174M12"

BASE COUNT 46879 a 41019 c 43482 g 52771 t  
ORIGIN  
Query Match 17.8%; Score 181.6; DB 10; Length 184151;  
Best Local Similarity 91.0%; Pred. No. 3.2e-38;  
Matches 193; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
Qy 593 AGGCTCTGAACACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAAAATGCCAGATGACTGCT 652  
Db 125770 AGGCCCTGAACACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAAAATGCCAGATGACTGCT 125711  
Qy 653 ACCAGTCCAAAGCTGAAGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAG 712  
Db 125710 ACCAGTCTAAACTGAAGGAGAACTGACCCCAATCTTTCATCTCCAGAAACCATAGCAG 125651  
Qy 713 ATGCTGTCAAAATCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGGACCTTTGTGGATG 772  
Db 125650 ATGCTGTCAAAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATTAAGAGACCTTTGTGGATG 125591  
Qy 773 ATATCTTCACTGTACAGAGGATGAATTAAG 804  
Db 125590 ATGTCTTCACTGTACAGGAGATGAATCAAG 125559

Search completed: June 24, 2003, 10:29:39  
Job time : 2674 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 04:05:10 ; Search time 1462 Seconds  
(without alignments)  
11277.022 Million cell updates/sec

Title: US-09-889-609B-1

Perfect score: 1018

Sequence: 1 atgtgtctcagctactgcat.....ttaccagacgtctgtttaa 1018

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154056 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	996	97.8	1398	11 AK017235	AK017235 Mus muscu
2	642.4	63.1	769	12 BF163794	BF163794 601769701
3	603.8	59.3	785	13 BF163753	BF163753 603047780
4	599.6	58.9	848	9 AU131397	AU131397 AU131397
5	595.8	58.5	667	10 BB609829	BB609829 BB609829
6	583.8	57.3	600	13 BF198879	BF198879 4022-49 M

7	554.2	54.4	565	10 BE655084	BE655084 UI-M-ALI-
8	543.8	53.4	819	13 BF1739030	BF1739030 603359880
9	496.2	48.7	918	14 BQ885621	BQ885621 AGENCOURT
10	479.8	47.1	823	13 BF1738966	BF1738966 603359780
11	475.2	46.7	698	9 AV022510	AV022510 AV022510
12	461.2	45.3	602	10 AV374491	AV374491 AV374491
13	460.4	45.2	634	14 BM1719814	BM1719814 UI-E-EJO-
14	455	44.7	455	9 AI322578	AI322578 m151e12.y
15	445.6	43.8	729	13 BG9565678	BG9565678 602830522
16	429.2	42.2	793	12 BG404240	BG404240 602420304
17	428.4	42.1	527	9 AA446793	AA446793 z889f02.r
18	418.6	41.1	827	13 BF1332919	BF1332919 602984289
19	406.4	39.9	556	10 BB692486	BB692486 BB692486
20	400.8	39.4	830	12 BF135412	BF135412 601784509
21	393.2	38.6	519	10 BB641631	BB641631 BB641631
22	393.2	38.6	652	10 AV709023	AV709023 AV709023
23	386.8	38.0	697	10 AV649313	AV649313 AV649313
24	367.2	36.1	688	10 BB625125	BB625125 BB625125
25	367	36.1	380	9 AA034539	AA034539 m151e12.r
26	366.2	36.0	661	10 BB621410	BB621410 BB621410
27	353.8	34.8	1087	12 BF532321	BF532321 602073350
28	334.4	32.8	645	10 AW211242	AW211242 u129q12.y
29	333.8	32.8	646	10 BB644584	BB644584 BB644584
30	325.4	32.0	391	14 BQ303565	BQ303565 RC5-BT074
31	323.8	31.8	1159	14 BQ215405	BQ215405 AGENCOURT
32	323	31.7	404	14 BQ303566	BQ303566 RC5-BT074
33	322	31.6	386	14 BQ303566	BQ303566 RC5-BT074
34	320.2	31.5	657	10 AW319255	AW319255 um8405.y
35	317.2	31.2	611	10 BB618702	BB618702 BB618702
36	316	31.0	632	10 BB641468	BB641468 BB641468
37	314.2	30.9	672	10 BE380996	BE380996 601271533
38	300.8	29.5	1045	12 BF384769	BF384769 602046261
39	284	27.9	471	14 H73097	H73097 yu25h09.r1
40	283.4	27.8	394	10 AV649514	AV649514 AV649514
41	269.8	26.5	742	9 A1525507	A1525507 PTL1.1_4_B
42	267.6	26.3	438	14 H86748	H86748 yt08h10.r1
43	265.6	26.1	572	10 BB651611	BB651611 BB651611
44	253	24.9	698	10 AW550801	AW550801 L0069E08-
45	253	24.9	698	13 BM213835	BM213835 C0842C08-

#### ALIGNMENTS

RESULT 1	AK017235	AK017235	1398 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	AK017235	Mus musculus adult male pituitary gland cDNA, RIKEN full-length enriched library, clone:5330405D10:serine racemase, full insert sequence.				
DEFINITION	AK017235	HTC; CAP trapper.				
ACCESSION	AK017235	1 GI:12856372				
VERSION	AK017235	HTC; CAP trapper.				
KEYWORDS	AK017235	Mus musculus (strain:C57BL/6J) adult male pituitary gland cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library				
SOURCE	AK017235	clone:5330405D10.				
ORGANISM	AK017235	Mus musculus				
REFERENCE	AK017235	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	AK017235	1 Carninci, P., and Hayashizaki, Y.				
TITLE	AK017235	High-efficiency full-length cDNA cloning				
JOURNAL	AK017235	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	AK017235	99279253				
PUBMED	AK017235	10349636				
REFERENCE	AK017235	2				
AUTHORS	AK017235	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	AK017235	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	AK017235	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	AK017235	20499374				
PUBMED	AK017235	11042159				

2526 88/1

May



Db 890 AAACCTAGTGTGAAGATATACGCTGCTGAGCCCTCGAATGCAGATGACTGCTACCAAGTCT 949  
QY 661 AAACCTAGTGTGAAGATATACGCTGCTGAGCCCTCGAATGCAGATGACTGCTACCAAGTCT 949  
Db 950 AAACCTAGTGTGAAGATATACGCTGCTGAGCCCTCGAATGCAGATGACTGCTACCAAGTCT 949  
QY 721 AAACCTAGTGTGAAGATATACGCTGCTGAGCCCTCGAATGCAGATGACTGCTACCAAGTCT 949  
Db 1010 AAACCTAGTGTGAAGATATACGCTGCTGAGCCCTCGAATGCAGATGACTGCTACCAAGTCT 949  
QY 781 ACTGTCCAGAGATGAATCAAGTATCAACCCAGCTGGTGGGGGAGAAATGAACATG 840  
Db 1070 ACTGTCCAGAGATGAATCAAGTATCAACCCAGCTGGTGGGGGAGAAATGAACATG 1129  
QY 841 CTCATTGAGCCGACTGCTGGCTGGCAGCTGGTGGCAGTCTGCTCAGCATTTCACAAACA 900  
Db 1130 CTCATTGAGCCGACTGCTGGCTGGCAGCTGGTGGCAGTCTGCTCAGCATTTCACAAACA 1189  
QY 901 GTCTCTCCAGAGTAAAGAACCTGTGCATTGTACTGAGTGGGGGAAATAGACCTAAC 960  
Db 1190 GTCTCTCCAGAGTAAAGAACCTGTGCATTGTACTGAGTGGGGGAAATAGACCTAAC 1249  
QY 961 TCCCTGAACCTGGTGGGGGAGCTGAACGGCCAGCTCCCTTACCAGACGGTCTGTT 1016  
Db 1250 TCCCTGAACCTGGTGGGGGAGCTGAACGGCCAGCTCCCTTACCAGACGGTCTGTT 1305

RESULT 2  
BF163794  
LOCUS 601769701F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3988873 5',  
DEFINITION mRNA sequence.  
ACCESSION BF163794  
VERSION BF163794.1 GI:11043971  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Arrayed by: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM9198 row: a column: 02  
High quality sequence stop: 699.

FEATURES  
source  
1. .769  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C2EHC II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3988873"  
/clone\_lib="NCI\_CGAP\_Lu29"  
/tissue\_type="spontaneous tumor, metastatic to mammary."  
stem cell origin.  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"  
BASE COUNT 225 a 186 c 187 g 171 t  
ORIGIN  
Query Match 63.1%; Score 642.4; DB 12; Length 769;  
Best Local Similarity 98.7%; Pred. No. 3e-183;

Matches 679; Conservative 0; Mismatches 6; Indels 3; Gaps 3;  
QY 1 ATGTGTGCTCAGTACTGCTATCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTC 60  
Db 40 ATGTGTGCTCAGTACTGCTATCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTC 99  
QY 61 GACTCTATCAGCTCACCAGTGTGCTTAAACAGCTCCATTTTGAATCAAAATAGCAGGCGC 120  
Db 100 GACTCTATCAGCTCACCAGTGTGCTTAAACAGCTCCATTTTGAATCAAAATAGCAGGCGC 159  
QY 121 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAAGCTGGGTCTTTTGAATTCAGGTGCC 180  
Db 160 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAAGCTGGGTCTTTTGAATTCAGGTGCC 219  
QY 181 CTTAATGCCATCAGAGGCTTAAATCTGACACGCC - AGAAGAGAACCCCAAGCCCTAGT 239  
Db 220 CTTAATGCCATCAGAGGCTTAAATCTGACACGCCCAAGAGAGAACCCCAAGCCCTAGT 279  
QY 240 TACTCACAGCAGCGAAACCATGGCCAGCTCTCACTATGTCTTAAACTGGAAGAAAT 299  
Db 280 TACTCACAGCAGCGAAACCATGGCCAGCTCTCACTATGTCTTAAACTGGAAGAAAT 339  
QY 300 TCTCTGCTTACATTTGGTTTCCCAACAGCTCCCAACTGCAAGAAACTGGCAATCCAAGC 359  
Db 340 TCTCTGCTTACATTTGGTTTCCCAACAGCTCCCAACTGCAAGAAACTGGCAATCCAAGC 399  
QY 360 CTATGGAGCATCGATAGTATCTGACCCCAAGTGCAGAGTCCAGAGAAAGTCACTCA 419  
Db 400 CTATGGAGCATCGATAGTATCTGACCCCAAGTGCAGAGTCCAGAGAAAGTCACTCA 459  
QY 420 AAGAATTATGCAAGAAACAGAGGATCTTGGTCCATCCCAACAGAGGCTTCACTGAT 479  
Db 460 AAGAATTATGCAAGAAACAGAGGATCTTGGTCCATCCCAACAGAGGCTTCACTGAT 519  
QY 480 AGCTGGACAGAGAACAAATTTGCCCTGGAAGTGTGACACAGGTTCCCTGGTAGATCACT 539  
Db 520 AGCTGGACAGAGAACAAATTTGCCCTGGAAGTGTGACACAGGTTCCCTGGTAGATCACT 579  
QY 540 GGTGTGACAGTACAGAGGAGGAGGATGTTGCTGGAATAGCCATTACCAATTAAGCCCT 599  
Db 580 GGTGTGACAGTACAGAGGAGGAGGATGTTGCTGGAATAGCCATTACCAATTAAGC - CCT 638  
QY 600 GAAACCTAGTGTGAAGTATACGCTGCTGAGCCCTCGAATGACAGTCTGCTTACCAGTC 659  
Db 639 GAAACCTAGTGTGAAGTATACGCTGCTGAGCCCTTGAATGC - GATGACTGGTACCAGTC 697  
QY 660 TAAACTGAAGAGAGACTGACCCCAAT 687  
Db 698 TAAACTGAAGAGAGAACTGGCCCAAT 725

RESULT 3  
BF163753  
LOCUS 603047780F1 NIH\_MGC\_116 Homo sapiens cDNA clone IMAGE:518099 5',  
DEFINITION mRNA sequence.  
ACCESSION BF163753  
VERSION BF163753.1 GI:15755331  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM11469 row: p column: 20

High quality sequence stop: 783.

Location/Qualifiers

#### FEATURES

source  
1. .785  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5188099"  
/clone\_lib="NIH\_MGC\_116"  
/lab\_host="DH10B"  
/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPT6; Site1: NotI; Site2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH\_MGC Library."  
BASE COUNT 225 a 168 c 200 g 192 t  
ORIGIN

Query Match 59.3%; Score 603.8; DB 13; Length 785;

Best Local Similarity 87.9%; Pred. No. 1.5e-171;

Matches 602; Conservative 0; Mismatches 92; Indels 3; Gaps 3;

QY 104 ATCAATAGCAGGGCCCAATCTTTCTTCAATGTGAGCTCTTCCAGAAACTGGGTCTT 163  
DB 1 ATCAACTAAGCAGGGCCCAATCTTTCTTCAATGTGAGCTCTTCCAGAAACTGGGTCTT 60  
QY 164 -TTAAGATTCGAGGTGCCCTTAATGCCATCAGAGGCTTAATCTCGACAGCCGAGAAG 222  
DB 61 GTTAAGATTCGAGGTGCCCTTAATGCCATCAGAGGCTTAATCTCGACAGCCGAGAAG 120  
QY 223 AAGCCCAAGCGGTAGTTACTCACAGCAGGAAACCATGGCCAGCTCTCACTATGCT 282  
DB 121 AAGCCGAAGCTGT-GTTACTCACAGCATGGAACCATGGCCAGCTCTCACTATGCT 179  
QY 283 GCTAACTGGAGGAATTCCTGCTTACATTTGGTTCCCAACAGCTCCCAACTGCAAG 342  
DB 180 GCCAAATGGAGGAATTCCTGCTTATATTGGTGGCCCGACAGAGCTCCAGACTGTA 239  
QY 343 AACTGGCAATCAAGCCTTAGGCATCGATAGTATATCTGTGACCAAGTCAAGATCC 402  
DB 240 AACTTGGCAATCAAGCCTTAGGCATCGATAGTATATCTGTGACCAAGTCAAGATCC 299  
QY 403 AGAGAAAGGTCACCTCAAGAAATTTATGCAAGAAACAGAGGATCTTGGTCCATCCCA 462  
DB 300 AGAGAAATGTTGCAAAAGAGTTACAGAGAAACAGAGGATCATGTTATATCCCAAC 359  
QY 463 CAGGAGCCTGCAAGTATAGCTGGACAGGAACAATTTGCCCTGGAAGTGTGTAACCAAG 522  
DB 360 CAGGAGCCTGCAAGTATAGCTGGACAGGAACAATTTGCCCTGGAAGTGTGTAACCAAG 419  
QY 523 CCCTTGGTAGATGCAGTGGTGTACCATGAGGAGGAGGAGTGGTCTGGAATAGCC 582  
DB 420 CCCTTGGTAGATGCAGTGGTGTACCATGAGGAGGAGGAGTGGTCTGGAATAGCA 479  
QY 583 ATTACAAATTAAGCCCTGAAACCTTAGTGTGAAGGTATAGGCTGTGAGCCCTCGAATG 642  
DB 480 ATTACAGTTAAGGCTCTGAAACCTTAGTGTGAAGGTATATGCTGTGAAACCTCGAATG 539  
QY 643 GATGACTGCTACAGTCTAAATGAAAGAGAGACTGACCCCAATCTTTCATCCTCCAGAA 702  
DB 540 GATGACTGCTACAGTCTAAATGAAAGAGAGACTGATGCCCAATCTTTCATCCTCCAGAA 599  
QY 703 ACCATAGCAGATGGTGTCAATTCACAGCATTTGCTTGAATACCTGGCTATTATAGAGAC 762  
DB 600 ACCATAGCAGATGGTGTCAATTCACAGCATTTGCTTGAATACCTGGCTATTATAGAGAC 659

QY 763 CTTGTGGATGATCTCTTCACTGTCACCGAAGATGAATCAAGTATCAACCCAGCTGGTG 822  
DB 660 CTTGTGGATGATCTCTTCACTGTCACAGAGATGAATTAAGTGTGC-ACCCAGCTGGTG 718  
QY 823 TGGGGGAGAAATGAACACTGCTCAATTGAGCGCACTGCTGGGCTGGCACTGGCTGCAGTCTG 882  
DB 719 TGGGAGAGGATGAACACTACTCATTTGAACCTACAGCTGGTGTGGAGCTGGCTGCTGTGCTG 778  
QY 883 TCTCAGC 889  
DB 779 TCTCAAC 785

#### RESULT 4

AU131397 LOCUS 848 bp mRNA linear EST 01-AUG-2002  
AU131397 NT2RP3 Homo sapiens cDNA clone NT2RP3002501 5', mRNA  
sequence.

#### ACCESSION

AU131397

#### VERSION

AU131397.1 GI:10991751

#### KEYWORDS

EST.

#### SOURCE

human.

#### ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

#### REFERENCE

AUTHORS

OTAI, F., NISHIKAWA, T., SUZUKI, Y., ISHII, S., SAITO, K., KAWAI, Y.,  
YAMAMOTO, J., WAKAMATSU, A., NAKAMURA, Y., NAGAI, T., SUGANO, S. and  
ISOGAI, T.

#### TITLE

HRI human cDNA project

#### JOURNAL

Unpublished (2000)

#### COMMENT

Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- and 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

#### FEATURES

Location/Qualifiers

1. 848

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="NT2RP3002501"

/clone\_lib="NT2RP3"

/cell\_type="teratocarcinoma"

/cell\_line="NT2"

/note="Vector: pME18SPL3; mRNA from NT2 neuronal precursor  
cells after 2-weeks retinoic acid (RA) induction"

BASE COUNT 236 a 187 c 214 g 207 t

#### ORIGIN

Query Match 58.9%; Score 599.6; DB 9; Length 848;

Best Local Similarity 87.6%; Pred. No. 3e-170;

Matches 687; Conservative 0; Mismatches 93; Indels 4; Gaps 3;

QY 1 ATGTGTGCTCAGTACTGCACTCTTCTTGTGATGTTGAAAAAGCTCATATCAACATTCAA 60  
DB 69 ATGTGTGCTCAGTATTGTCATCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCAA 128  
QY 61 GACTCTATCCACCTCACCCAGTGTGTAAACAGCTCCATTTTGAATCAATAGCAGGCCG 120  
DB 129 GATTCTTCACTCACTCACACCACTGTGTAAACAGCTCCATTTTGAATCAATAGCAGGCCG 188  
QY 121 ATCTTTCTTCAATATGTGAGCTCTTCCAGAAAACCTGGTCTTTTAAGATTTCAGGTGCC 180  
DB 189 AATCTTTCTTCAATATGTGAGCTCTTCCAGAAAACAGGATCTTTTAAGATTTCAGGTGCC 248  
QY 181 CTTAATGCCATCAGAGGCTTAATCTCTGACGCCGAGAGAGCCCAAGCCGCTAGTT 240

```

Db      249 CTCAATGCCGTGAGAGCTTGGTCTCTGATGCTTTAGAAAGGAGCCGAAAGCTCTGTGT 308
QY      241 ACTCAGCAGCGGAAACCATGCGCAAGCTCTCACCCTAGCTGCTTAAACTGCAAGGAATT 300
Db      309 ACTCAGCAGGTGGAAACCATGCGCAGGCTCTCACCCTAGCTGCTTAAACTGCAAGGAATT 368
QY      301 CTGCTTACATTTGTTGTTCCCAACAGCTCCCAACTGCAAGAACTGGCAATCAAGCC 360
Db      369 CTGCTTATATTTGTTGTTCCCAACAGCTCCCAACTGCAAGAACTGGCAATCAAGCC 428
QY      361 FATGAGCATCATAGTATGCTGACCAAGTACAGAGTACAGAGAAAGTCTACTCAA 420
Db      429 TACGGAGCGTCAATTTGATCTGTGAACCTAGTGTAGTCCAGAGAAATCTTGCAGAA 488
QY      421 AGAATTTATGCAAGAAACAGGCACTTTGTCCTCATCCCAACCCAGGAGCTCCAGTGATA 480
Db      489 AGAGTTACAGAAAGCAAGAGGCAATCATGGTACATCCCAACCCAGGAGCTCCAGTGATA 548
QY      481 GCTGACAAGGAACAATTTGCCCTGGAAGTGTCTGAACCGAGTTCCCTTGGTAGATCACTG 540
Db      549 GCTGACAAGGACAAATTTGCCCTGGAAGTGTCTGAACCGAGTTCCCTTGGTAGATCACTG 608
QY      541 GTGGTACCATGATGAGGAGGAGGAATGTTGCTGGAATAGCCATACAAATTAAGGCCCTG 600
Db      609 GTGGTACCTGTAGGTGGAGAAATGCTTGTGGAATAGCAATTAACAGTTAAGGCTCTG 668
QY      601 AAACCTAGTGTGAAGTATACCTGCTGAGCCCTCGAATGCAGATGACTGTACCACT 660
Db      669 AAACCTAGTGTGAAGTATATCTGCTGAACCTCAATATGCAGATGACTGTACCACT 727
QY      661 AAACCTGAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGCTGTC 720
Db      728 AAGCTGAAGGAGAACTGATGCCCAATCTTATCTCTCAGAA-CCATACAGATGCTGTC 785
QY      721 AAATCAGCATTTGGCTTGAATACCTGGCTTATTAAGAGACCTTGTGGATGATGCTTTC 780
Db      786 AAATCCA-CATTGGCTTGAACACCTGGGCTTATATCANGGACCTTGGNGATGATATCTTT 844
QY      781 ACTG 784
Db      845 ACTG 848

```

## RESULT 5

BB609829

LOCUS

BB609829 BB609829 RIKEN full-length enriched, 18 days embryo Mus musculus  
cDNA clone 1190028F16 5', mRNA sequence.

ACCESSION

BB609829

VERSION

BB609829.1

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.

RIKEN House ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

JOURNAL

COMMENT

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

## FEATURES

source

1..867  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="RIKEN full-length enriched, 18 days embryo"  
/dev\_stage="18 days embryo"  
/lab\_host="SOLR"  
/note="Site\_1: XhoI; Site\_2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st-strand cDNA was primed with a primer [5'  
GAGAGAGAGCGCGCGCAACTCGAGTGTGTTTATTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'  
GAGAGAGAGAGGATCCAGAGCTCAATTAATTAATTAACCCCCCCCC 3'].  
cDNA was cleaved with XhoI and SstI."

BASE COUNT

ORIGIN

190 a 163 c 147 t 1 others

Query Match

Best Local Similarity

Matches 597; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTGTCCTCAGTACTGCTATCTCTTGTGATGTTGAAAAGCTCATATCAACATTCAA 60

Db 68 ATGTGTCCTCAGTACTGCTATCTCTTGTGATGTTGAAAAGCTCATATCAACATTCAA 127

QY 61 GACTCTATCCACCTCACCCTGCTGTCAACAGCTCATTTTGAATCAATAGCAGGCGC 120

Db 128 GACTCTATCCACCTCACCCTGCTGTCAACAGCTCATTTTGAATCAATAGCAGGCGC 187

QY 121 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAACTGGCTCTTTTAAGATTCGAGTGCC 180

Db 188 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAACTGGCTCTTTTAAGATTCGAGTGCC 247

QY 181 CTTAATGCCATCAGAGCTTAAATCTTGACAGCCAGAGAGAGCCCAAGCCGTAGTT 240

Db 248 CTTAATGCCATCAGAGCTTAAATCTTGACAGCCAGAGAGAGCCCAAGCCGTAGTT 307

QY 241 ACTCACAGCAGCGGAAACCACTGGCCAACTCTCACCTATGCTGTAACTGGAAGGAATT 300

```

Db      308 ACTCAGCAGCGGAAACCATGCGCAAGCTCTCACCTATGCTGCTAAACTGGAAGAAAT 367
QY      301 CTCTCTATCCACTCACCCTACCCAGTCTTAAAGCTTCAATTTTGAATCAATAGCAGGCGC 120
Db      366 GACTNTATCACCTCACCCTACCCAGTCTTAAAGCTTCAATTTTGAATCAATAGCAGGCGC 125
QY      121 AATCTTTTCTTCAATTTGAGCTCTTCCAGAAACTGGGTCTTTTAAAGTTTCCAGGTGCC 180
Db      126 AATCTTTTCTTCAATTTGAGCTCTTCCAGAAACTGGGTCTTTTAAAGTTTCCAGGTGCC 185
QY      181 CTTAATGCCATCAGAGGCTTAAATTTCTGACGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db      186 CTTAATGCCATCAGAGGCTTAAATTTCTGACGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 245
QY      241 ACTCAGCAGCGGAAACCATGCGCAAGCTCTTCAACCTATGCTGCTTAACTGGAAGAAAT 300
Db      246 ACTCAGCAGCGGAAACCATGCGCAAGCTCTTCAACCTATGCTGCTTAACTGGAAGAAAT 305
QY      301 CTCTCTTACATTTGCTTCCCAACAGCTCCCAACTGCAAGAACTGCAATCCCAAGCC 360
Db      306 CTCTCTTACATTTGCTTCCCAACAGCTCCCAACTGCAAGAACTGCAATCCCAAGCC 365
QY      361 TATGAGCAGTCGATAGTACTACTGTGACCCCAAGTGAGAGTCCAGAGAGAGAGAGAGAGAG 420
Db      366 TATGAGCAGTCGATAGTACTACTGTGACCCCAAGTGAGAGTCCAGAGAGAGAGAGAGAGAG 425
QY      421 AGAATTATCAAGAAACAGAGCATCTTGGTCCATCCCAAGTGAGAGTCCCAAGTGAGAGAGAG 480
Db      426 AGAATTATCAAGAAACAGAGCATCTTGGTCCATCCCAAGTGAGAGTCCCAAGTGAGAGAGAG 485
QY      481 GCTGGACAAAGAACAAATTTGCCCTGGAAAGTCTGCAAGAGTCTGCTGCTGCTGCTGCTGCT 540
Db      486 GCTGGACAAAGAACAAATTTGCCCTGGAAAGTCTGCAAGAGTCTGCTGCTGCTGCTGCTGCT 545
QY      541 GTGTGTACCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 595
Db      546 GTGTGTNCCANTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600

RESULT 7
LOCUS   BE655084
DEFINITION 565 bp mRNA linear EST 06-SEP-2000
ACCESSION UI-M-ALL1-ahr-f-12-0-UI-r2 NIH_BMAP_MCO_N Mus musculus cDNA clone
VERSION    BE655084
KEYWORDS   BE655084.1 GI:9980997
SOURCE     house mouse
ORGANISM   Mus musculus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      1 (bases 1 to 565)
JOURNAL    Bonaldo,M.F., Lennon,G. and Soares,M.B.
MEDLINE    Normalization and subtraction: two approaches to facilitate gene
COMMENT    discovery
           Genome Res. 6 (9), 791-806 (1996)
           9704477
           Contact: Chin, H
           National Institute of Mental Health
           6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
           20892-9643, USA
           Tel: 301 443 1706
           Fax: 301 443 9890
           Email: mEST@mail.nih.gov
           CDNA Library Preparation: M.B. Soares Lab Clone distribution:
           Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
           should be noted that Bento Soares is generating a small number of
           additional specialized non-redundant arrays of BMAP cDNAs whose
           availability will be considered under appropriate and limited
           collaborative arrangements
           Seq primer: M13 Reverse.
           Location/Qualifiers
           1. .565

BASE COUNT 177 a 145 c 134 g 136 t 8 others
ORIGIN
Query Match 57.3%; Score 583.8; DB 13; Length 600;
Best Local Similarity 98.3%; Pred. No. 1.5e-165;
Matches 585; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY      1 ATGTGTGCTCAGTACTGCTATCCCTTCTGCTGATGTTGAAAGAGCTCATATCAACATTCAA 60

```

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AL1-f-12-0-UI"
/clone_lib="NIH_BMAP_MCO_N"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_MCO_N library is a normalized library constructed
from mouse cortex. The tag is a string of 5 nucleotides
present between the Not I site and the oligo-dT track.
The library was constructed as described by Bonaldo,
Lennon and Soares, Genome Research 6: 791-806, 1996.
Tissue provided by Ms. Annie Novakovich, Zivic-Miller
Laboratories."
BASE COUNT      152 a 132 c 152 g 129 t
ORIGIN

Query Match      54.4%; Score 554.2; DB 10; Length 565;
Best Local Similarity 99.5%; Pred. No. 1.4e-156;
Matches 556; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 458 CCAACAGAGCCTGCAGTGTAGCTGGACAGGACAAATTCCTCGGAAGTCTGAACC 517
DB 1 CCAACAGAGGCTGCAGTGTAGCTGGACAGGACAAATTCCTCGGAAGTCTGAACC 60
QY 518 AGGTTCCCTTGTGTAGATGCACGTGGTGTACCAAGTAGGAGGAGGAATGGTGTGGA 577
DB 61 AGGTTCCCTTGTGTAGATGCACGTGGTGTACCAAGTAGGAGGAGGAATGGTGTGGA 120
QY 578 TAGCCATTACAATTAAAGGCCCTGAAACCTAGTGTGAAGGTATACGCTGTCGAGCCCTCGA 637
DB 121 TAGCCATTACAATTAAAGGCCCTGAAACCTAGTGTGAAGGTATACGCTGTCGAGCCCTCGA 180
QY 638 ATGCAGATGACGTCTACCACTCTAACTCAAGGAGAACTGACCCCAATCTTCATCTC 597
DB 181 ATGCAGATGACGTCTACCACTCTAACTCAAGGAGAACTGACCCCAATCTTCATCTC 240
QY 698 CAGAAACCATAGCAGATGGTGTCAATCCAGCATTGGCTTGAATACCTGGCCCTATTATAA 757
DB 241 CAGAAACCATAGCAGATGGTGTCAATCCAGCATTGGCTTGAATACCTGGCCCTATTATAA 300
QY 758 GAGACCTTGTGATGATGCTTTCACTGTCCAGAGATGAATCAAGTATGCAACCCAGC 817
DB 301 GAGACCTTGTGATGATGCTTTCACTGTCCAGAGATGAATCAAGTATGCAACCCAGC 360
QY 818 TGGTGTGGGGAGATGAACCTGCTCATGACCCGACTGCTGGCGTGGCAGTGGCTGCAG 877
DB 361 TGGTGTGGGGAGATGAACCTGCTCATGACCCGACTGCTGGCGTGGCAGTGGCTGCAG 420
QY 878 TGCTGTCTCAGCATTTCCAAACAGTCTCTCCAGAAAGTAAAGAACGCTCTCATTTGTACTCA 937
DB 421 TGCTGTCTCAGCATTTCCAAACAGTCTCTCCAGAAAGTAAAGAACGCTCTCATTTGTACTCA 480
QY 938 GTGGGGGGAATGTAGACCTTAACCTCCCTGAATGGGTGGGGGAGGCTGAACGCCAGCTC 997
DB 481 GTGGGGGGAATGTAGACCTTAACCTCCCTGAATGGGTGGGGGAGGCTGAACGCCAGCTC 540
QY 998 CTTACCAGACGCTCTGTTT 1016
DB 541 CTTACCAGACGCTCTGTTT 559

RESULT 8
LOCUS B1739030
DEFINITION B1739030 819 bp mRNA linear EST 20-SEP-2001
603359880F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5367042 5',
mRNA sequence.
ACCESSION B1739030
VERSION B1739030.1 GI:15716056
KEYWORDS EST.
SOURCE house mouse.

```

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue procurement: The Cepko Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM11933 row: p column: 19  
High quality sequence stop: 817.

Location/Qualifiers  
1. 819  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5367042"  
/clone\_lib="NIH\_MGC\_94"  
/tissue\_type="retina"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 3.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
BASE COUNT 236 a 208 c 192 g 183 t  
ORIGIN

Query Match 53.4%; Score 543.8; DB 13; Length 819;  
Best Local Similarity 99.6%; Pred. No. 2.4e-153;  
Matches 545; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGTGTGCTCAGTACTGCTATCCTTTGCTGTGTTGAAAAAGCTCATATCAACATTCAA 60  
DB 273 ATGTGTGCTCAGTACTGCTATCCTTTGCTGTGTTGAAAAAGCTCATATCAACATTCAA 332  
QY 61 GACTCTATCCACTCACCCTCAGTGTACAAAGCTCCATTTTGAATCAAAATAGCAGGCGC 120  
DB 333 GACTCTATCCACTCACCCTCAGTGTACAAAGCTCCATTTTGAATCAAAATAGCAGGCGC 392  
QY 121 AATCTTTTCTTCAATGTGAGCTCTCCAGAAACTGGGCTCTTTTAAAGATTTCAGGTGCC 180  
DB 393 AATCTTTTCTTCAATGTGAGCTCTCCAGAAACTGGGCTCTTTTAAAGATTTCAGGTGCC 452  
QY 181 CTTAATGCCATCAGAGGCTTAATCTCTGACAGCCAGAGAGAGCCCAAGCCGTAGTT 240  
DB 453 CTTAATGCCATCAGAGGCTTAATCTCTGACAGCCAGAGAGAGAGCCCAAGCCGTAGTT 512  
QY 241 ACTCAGAGCGGGAACCATGCCCCAGCTCTCACCTATGCTGCTAACTGGAAGGAATT 300  
DB 513 ACTCAGAGCGGGAACCATGCCCCAGCTCTCACCTATGCTGCTAACTGGAAGGAATT 572  
QY 301 CTTGCTTACATTGTGGTCTCCCAACAGCTCCCAACTGCAAGAACTGCAATCCAGCC 360  
DB 573 CTTGCTTACATTGTGGTCTCCCAACAGCTCCCAACTGCAAGAACTGCAATCCAGCC 632  
QY 361 TATGGAGCATCGATAGTATATCTGTACCCCAAGTGCAGAGTCCAGAGAAAAGGTCACTCAA 420  
DB 633 TATGGAGCATCGATAGTATATCTGTACCCCAAGTGCAGAGTCCAGAGAAAAGGTCACTCAA 692  
QY 421 AGAATTATCAGAGAAACAGAGAGCATCTTGGTCCATCCCAACAGAGAGCTCGAGTGATA 480  
DB 693 AGAATTATCAGAGAAACAGAGAGCATCTTGGTCCATCCCAACAGAGAGCTCGAGTGATA 752  
QY 481 GCTGGACAGGACAAATTCCTCCCTGGAAGTGTGTAACCAAGGTTCCCTTGGTAGATGACATG 540  
DB 753 GCTGGACAGGACAAATTCCTCCCTGGAAGTGTGTAACCAAGGTTCCCTTGGTAGATGACATG 812

```

- QY      541 GTGGTAC 547
      Db      813 GTGGTAC 819

RESULT 9
BQ885621
LOCUS      918 bp      mRNA      linear      EST 16-AUG-2002
DEFINITION AGENCOURT_8754012 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332599
5', mRNA sequence.
ACCESSION BQ885621
VERSION   BQ885621.1 GI:22277639
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
           1 (bases 1 to 918)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
           Ph.D.
           cDNA Library Preparation: ResGen, Invitrogen Corp
           DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM13790 row: h column: 08
           High quality sequence stop: 609.

FEATURES             Location/Qualifiers
     source            1..918
                       /organism="Mus musculus"
                       /db_xref="taxon:10090"
                       /clone="IMAGE:6332599"
                       /clone_lib="NIH_MGC_130"
                       /lab_host="DH10B (phage-resistant)"
                       /note="Organ: oocytes; Vector: pCMV-SPORT6.1.cdbb;
                       Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
                       Primer: Oligo dt. Average insert size 1.95 kb.
                       Constructed by ResGen, Invitrogen Corp. Note: this is a
                       NIH_MGC Library."
BASE COUNT      247 a 230 c 230 g 203 t 8 others
ORIGIN

Query Match      48.7%; Score 496.2; DB 14; Length 918;
Best Local Similarity 93.3%; Pred. No. 7.1e-139;
Matches 571; Conservative 0; Mismatches 33; Indels 8; Gaps 5;

QY      1 ATGTGTCTCAGTACTGCTATCTCTTGTCTGATGTTGAAAACCTCATATCAACATCAA 60
      Db      308 ATGTGTCTCAGTACTGCTATCTCTTGTCTGATGTTGAAAACCTCATATCAACATCAA 367
      QY      61 GACTCTATCACTACCTCACCAGTGTCTTAACAAGCTCCATTTTGAATCAATAGCAGGCGC 120
      Db      368 GACTCTATCACTACCTCACCAGTGTCTTAACAAGCTCCATTTTGAATCAATAGCAGGCGC 427
      QY      121 ATCTTTTCTTCAATGTAGCTCTTCCAGAACTGGTCTTTTAAAGATTTCAGGTGCC 180
      Db      428 AATCTTTTCTTCAATGTAGCTCTTCCAGAACTGGTCTTTTAAAGATTTCAGGTGCC 487
      QY      181 CTTAATGCCATCAGAGGCTTAATCTCTGACAGCCAGAGAGAGCCCAAGCCGTAGTT 240
      Db      488 CTTAATGCCATCAGAGGCTTAATCTCTGACAGCCAGAGAGAGCCCAAGCCGTAGTT 547
      QY      241 ACTCACAGCAGGGAACCACTGCCAAGCTCTCACCCTATGCTGCTAACTGGAAGCAATT 300
      Db      548 ACTCACAGCAGGGAACCACTGCCAAGCTCTCACCCTATGCTGCTAACTGGAAGCAATT 607

```

---

```

QY      301 CTGTGTTACATTGTGGTTCCTCCCAACAGCTCCCAACTGCAAGAACTGCAATTCAGGCC 360
      Db      608 CTGTGTTACATTGTGGTTCCTCCCAACAGCTCCCAACTGCAAGAACTGCAATTCAGGCC 667
      QY      361 TATGGAGCATCGATAGTACTGTGACCCCAAGTGCAGGAGTCCAGAGAAAGTCACTCAA 420
      Db      668 TATGGAGCATCGATAGTACTGTGACCCCAAGTGCAGGAGTCCAGAG-AAAGGTCACTCNA 726
      QY      421 AGAATTATCAAGAAACAGAGGCGATCTTGGTCCATCCCAACAGGAGGCTGCAGTGATA 480
      Db      727 AGAATTATTCAGAGACAGAGGCGATCTTGGTCCCATCCCAACAGGAGGCTGCAGTGATA 786
      QY      481 GTGGACAAAGGAACAATTCCTCCTGGAAGTCTGAACAGGTTCCCTTGG-TAGATGCACT 539
      Db      787 GCTGGACNAGGAACAATTCCTCCTGGAAGTCTGAACAGGTTCCCTTGGNTAGATGCACT 846
      QY      540 -GGTGGTACCAGTAGG--AGGAGGAGGAGTGGTGTCTGGAATA--GCCATTACATTTAA 593
      Db      847 GGTGGTACCAGTAGGAGGAGGAGGAGTGGTGTCTGGAATAAGCCATTTACATTTAA 906
      QY      594 GGCCCTGAAACC 605
      Db      907 GGCCCTGGAACC 918

RESULT 10
B1738966
LOCUS      823 bp      mRNA      linear      EST 20-SEP-2001
DEFINITION 603359780F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5367043 5',
mRNA sequence.
ACCESSION B1738966
VERSION   B1738966.1 GI:15715979
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
           NIH-MGC http://mgi.nci.nih.gov/.
           National Institutes of Health, Mammalian Gene Collection (MGC)
           Unpublished (1999)
           Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: The Cepko Laboratory
           cDNA Library Preparation: Life Technologies, Inc.
           DNA Sequencing by: Incyte Genomics, Inc.
           DNA Distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11933 row: p column: 20
           High quality sequence stop: 799.

FEATURES             Location/Qualifiers
     source            1..823
                       /organism="Mus musculus"
                       /db_xref="taxon:10090"
                       /clone="IMAGE:5367043"
                       /clone_lib="NIH_MGC_94"
                       /tissue_type="retina"
                       /lab_host="DH10B (phage-resistant)"
                       /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                       Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                       Average insert size 3.3 kb. Library enriched for
                       full-length clones and constructed by Life Technologies.
                       Note: this is a NIH_MGC Library."
BASE COUNT      237 a 208 c 193 g 184 t
ORIGIN

Query Match      47.1%; Score 479.8; DB 13; Length 823;
Best Local Similarity 98.9%; Pred. No. 6.2e-134;
Matches 525; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY      1 ATGTGTCTCAGTACTGCTATCTCTTGTCTGATGTTGAAAAGCTCATATCAACATCAA 60

```

```

Db      274 ATGTGCTCAGTACTCATCTCTCTTGGTGTGTTGAAAAGCTCATATCAACATTCAA 333
Qy      61  GACTCTATCCACCTCACCCAGTCTAAACAAGCTCCATTTTGAATCAAAATAGCAGGCGC 120
Db      334 GACTCTATCCACCTCACCCAGTCTAAACAAGCTCCATTTTGAATCAAAATAGCAGGCGC 393
Qy      121 AATCTTTTCTTCAAAATGTAGCTTCTCCAGAAACTGGGCTTTTAAAGATTCGAGGTGCC 180
Db      394 AATCTTTTCTTCAAAATGTAGCTTCTCCAGAAACTGGGCTTTTAAAGATTCGAGGTGCC 453
Qy      181 CTTAATGCCATCAGAGCTTAATCTCTGACACGCCAGAGAGAGCCCAAGCCGTAGTT 240
Db      454 CTTAATGCCATCAGAGCTTAATCTCTGACACGCCAGAGAGAGCCCAAGCCGTAGTT 513
Qy      241 ACTCACAGCAGCGGAACCATG-GCCAAAGCTCTCACCTATGCTGCTAACTGGAAGGAAT 299
Db      514 ACTCACAGCAGCGGAACCATGNGCCAAAGCTCTCACCTATGCTGCTAACTGGAAGGAAT 573
Qy      300 TCTGCTTACATGTGTGTTCCCAACAGCTCCCACTGCAAG-AAACTGGCAATCCAAAG 358
Db      574 TCTGCTTACATGTGTGTTCCCAACAGCTCCCACTGCAAGAAACTGGCAATCCAAAG 633
Qy      359 CCTATGAGCATCGATAGTATAC-TGTGACCCCAAGTGCACAGTCCAGAGAAAGGTCACT 417
Db      634 CCTATGAGCATCGATAGTATAC-TGTGACCCCAAGTGCACAGTCCAGAGAAAGGTCACT 693
Qy      418 CAAAGAAATTTATGCAAGAACAGAGGAGCATCTTGTGCTCATCCCAACAGGAGCCTGCACT 477
Db      594 CAAAGAAATTTATG-C-AGAACAGAGGAGCATCTTGTGCTCATCCCAACAGGAGCCTGCACT 752
Qy      478 ATAGCTGGACAGAGAACAAATGGCCCTGGAAGTGTGTAACCAAGTTCCTCTTG 528
Db      753 ATAGCTGGACAGAGAACAAATGGCCCTGGAAGTGTGTAACCAAGTTCCTCTTG 803

```

```

RESULT 11
AV022510      698 bp  mRNA  linear  EST 23-OCT-2001
LOCUS      AV022510 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA
DEFINITION  clone 1190028F16, mRNA sequence.
ACCESSION  AV022510
VERSION    AV022510.2 GI:16356391
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus

```

```

REFERENCE    Eukaryota: Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              1 (bases 1 to 698)

```

```

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)

```

```

On May 11, 1999 this sequence version replaced gi:4799502.
Contact: Chie Owa
Genome Science Laboratory

```

```

RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098

```

```

Email: genome-res@tc.riken.go.jp
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura

```

S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
 , Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa  
 K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
 Hayashizaki,Y.  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
 further details.  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.

#### FEATURES

Location/Qualifiers  
 1..698  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="1190028F16"  
 /clone\_lib="Mus musculus 18-day embryo C57BL/6J"  
 /sex="mixed"  
 /dev\_stage="18-day embryo"  
 /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia  
 ) with a modified polylinker; Site.1: Not I; Site.2: Eco  
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
 primer [5',  
 TGTACCAATCTGAATGGGAGCGCGCGAATGGTTTTTTTTTTTTTTTTTTTTTTTTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT7T3 vector.  
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library.  
 constructed and normalized by Bento Soares and M.Fatima  
 Bonaldo."

BASE COUNT 203 a 150 c 171 g 174 t

ORIGIN  
 Query Match 46.7%; Score 475.2; DB 9; Length 698;  
 Best Local Similarity 98.4%; Pred. No. 1.4e-132;  
 Matches 480; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 Qy 529 GTAGATGCACCTGGTGTACCAAGTAGGAGGAGGAGGAGTGGTGTGTTGGAATAGCCATTACA 588  
 Db 1 GTAGATGCACCTGGTGTACCAAGTAGGAGGAGGAGGAGTGGTGTGTTGGAATAGCCATTACA 60  
 Qy 589 ATTAAGGCCCTGAAACCTAGTGTGAAGGTATACGCTGTGAGCCCTCGAATGCGAGATGAC 648  
 Db 61 TTTAAGGCCCTGAAACCTAGTGTGAAGGTATACGCTGTGAGCCCTCGAATGCGAGATGAC 120  
 Qy 649 TGCTACCACTTAAACTGAAAGGAGAACTGACCCCAATCTTCATCTCCCAAGAACCATTA 708  
 Db 121 TGCTACCACTTAAACTGAAAGGAGAACTGACCCCAATCTTCATCTCCCAAGAACCATTA 180  
 Qy 709 GCAGATGGTGTCAAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATAAGAGACCTTGTG 768  
 Db 181 GCAGATGGTGTCAAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATAAGAGACCTTGTG 240  
 Qy 769 GATGATGCTTCTACTGTCCACGAAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGG 828  
 Db 241 GATGATGCTTCTACTGTCCACGAAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGG 300  
 Qy 829 AGAATGAACCTGCTCATTGAGCCGACCTGCTGGCGGTGGCAGCTGGCTGCTCTCAG 888  
 Db 301 AGAATGAACCTGCTCATTGAGCCGACCTGCTGGCGGTGGCAGCTGGCTGCTCTCAG 360

Qy 889 CATTTCACAAAGCTCTCTCCAGAAAGTAAGAACCTGCTGCAATGTACTCAGTGGGGGAAT 948  
 Db 361 CATTTCACAAAGCTCTCTCCAGAAAGTAAGAACCTGCTGCAATGTACTCAGTGGGGGAAT 420  
 Qy 949 GTACACCTAACCTCCCTGAACTGGTGGGCGAGCGCTGAAGCGGCAGCTCTTACCAGACG 1008  
 Db 421 GTACACCTAACCTCCCTGAACTGGTGGGCGAGCGCTGAAGCGGCAGCTCTTACCAGACG 480  
 Qy 1009 GTCGTGTTT 1016  
 Db 481 GTTCTGT 488

## RESULT 12

AV374491

## LOCUS

AV374491 602 bp mRNA linear EST 24-OCT-2001  
 cDNA clone 9130011P22 3', mRNA sequence.

## ACCESSION

AV374491

## VERSION

AV374491.2

## KEYWORDS

EST.

## SOURCE

house mouse.

## ORGANISM

Mus musculus

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 602)

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,

Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda

M., Koyu, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,

Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki

D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,

Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,

Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

On Nov 14, 1999 this sequence version replaced gi:6422138.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura

S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and

Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara

X. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa

K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and

Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp/) for

further details.

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

## FEATURES

## source

## prepare mouse tissues.

## Location/Qualifiers

1..602

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone\_lib="9130011P22"

/clone\_lib="RIKEN full-length enriched, adult male cecum"

/sex="male"

/tissue\_type="cecum"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Site 1: SalI; Site 2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'

GAGAGAGAGAAGGATCCAGAGCTCTTTTTTTTTTTTNN 3', cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by

cap-trapper. cDNA went through one round of normalization

to Rot = 10.0 and subtraction to Rot = 185.0. Second

strand cDNA was prepared with the primer adapter of

sequence [5' GAGAGAGAGATCTCGAGTTAATAATTAATCCCCCCCCCC

3']. cDNA was cloned into the XhoI and BamHI sites.

Vector: a modified pBluescript KS(+) after bulk excision

from lambda FLC I. Cloning sites, 5' end: SalI; 3' end:

BamHI"

## BASE COUNT

167 a 133 c 148 g 152 t

## ORIGIN

2 others.

## Query Match

45.3%; Score 461.2; DB 10; Length 602;

Best Local Similarity 94.7%; Pred. No. 2.2e-128;

Matches 487; Conservative 0; Mismatches 25; Indels 1;

Gy 503 TGGAGTGTGAACAGGTTCCCTTGATGATGACCTGGTGGTACCATAGGAGGAGAG 562

Db 1 TTGAAGTGTGAACAGGTTCCCTTGATGATGACCTGGTGGTACCATAGGAGGAGAG 58

Gy 563 GAATGGTGTGGAATAGCCATTACAAATTAAGGCCCTTGAACCTAGTGTGAAGGTATACG 622

Db 59 GAATGGTGTGGAATAGCCATTACAAATTAAGGCCCTTGAACCTAGTGTGAAGGTATACG 118

Gy 623 CTGCTGAGCCTCGAATGACAGTCTGCTACAGTCTAACTGAAAGGAGAGTGCACC 682

Db 119 CTGCTGAGCCTCGAATGACAGTCTGCTACAGTCTAACTGAAAGGAGAGTGCACC 178

Gy 683 CCAATCTTCATCTCCAGAAACCATAGCAGATGGTGTCAATCCAGCATTTGGCTTGAATA 742

Db 179 CCAATCTTCATCTCCAGAAACCATAGCAGATGGTGTCAATCCAGCATTTGGCTTGAATA 238

Gy 743 CCTGGCCTATTATAAGAGACCTTTGTGGATGATGTCTTCACTGTCCACGAAGATGAAATCA 802

Db 239 CCTGGCCTATTATAAGAGACCTTTGTGGATGATGTCTTCACTGTCCACGAAGATGAAATCA 298

Gy 803 AGTATGCAACCCAGCTGGTGTGGGGGAGATGAACCTGCTCATTTGAGCCGACCTGCTGGCG 862

Db 299 AGTATGCAACCCAGCTGGTGTGGGGGAGATGAACCTGCTCATTTGAGCCGACCTGCTGGCG 358

Gy 863 TGGCAGTGGCTGAGTGTCTGCTCAGCATTTTCAAAACAGTCTCTCCAGAGTAAAGAACG 922

Db 359 TGGCAGTGGCTGAGTGTCTGCTCAGCATTTTCAAAACAGTCTCTCCAGAGTAAAGAACG 418

Gy 923 TCTGCATTTGACTCAGTGGGGGAGATGAGACCTTAACCTCCCTGAACTGGGTGGGGCAGG 982

Db 419 TCTGCATTTGACTCAGTGGGGGAGATGAGACCTTAACCTCCCTGAACTGGGTGGGGCAGG 478

Gy 983 CTGAAGGGCGAGCTCTTACCACAGCGTCTGTTT 1016

Db 479 CTGAAGGGCGAGCTCTTACCACAGCGTCTGTTT 512

```

RESULT 13
BM719814
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

BM719814 634 bp mRNA linear EST 01-MAR-2002
UI-E-EJ0-ahu-j-16-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone
UI-E-EJ0-ahu-j-16-0-UI 5', mRNA sequence.
BM719814
BM719814.1 GI:19038647
human.
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 634)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.
Location/Qualifiers
1..634
/organism="Homo sapiens"
/lab_xref="taxon:9606"
/clone="UI-E-EJ0-ahu-j-16-0-UI"
/clone_lib="UI-E-EJ0"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/noted="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-EJ0 is a subcloned cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dr primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (df)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTACCGA; eye anterior segment, AATGCCCGCAT;
optic nerve, CCATTAAAGT; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
BASE COUNT 182 a 125 c 163 g 163 t
ORIGIN
Query Match 45.2%; Score 460.4; DB 14; Length 634;
Best Local Similarity 86.7%; Pred. No. 4e-128;
Matches 530; Conservative 0; Mismatches 77; Indels 4; Gaps 2;

QY 409 AAGGTCACTCAAGAATTATGCAAGAACAGAGGATCTTGGTCATCCCAACACGAGAG 468
|||||
1 AATGTTGCAAAAGAGTTACAGACAGAACAGAGGATCATGTGTACATCCCAACACGAGAG 50
|||||

QY 469 CCTGCAGTGATAGCTGGACAGGAACAATTCGCCCTGGAAGTCTGCAACCAAGTCCCTTG 528
|||||
61 CCTGCAGTGATAGCTGGACAGGACAATTCGCCCTGGAAGTCTGCAACCAAGTCCCTTG 120
|||||
529 GTAGATGCACCTGGTGTACCACTAGGAGGAGGAGGAATGTTGCTGGAATAGCAATTACA 588
|||||
121 GTGGATGCACCTGGTGTACCTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATTACA 180
|||||
589 ATTAAGGCCCTGAAACCTAGTGTGAAGGTATACGGTGTGAGCCCTCGAATCGAGATGAC 648
|||||
181 GTTAAGGCTCTGAAACCTAGTGTGAAGGTATATGCTGTGAACCCCTCAATCGAGATGAC 240
|||||
649 TGCTACCACTGTAAACTGAAAGAGAACTGACCCCAATCTTCATCTCCAGAAACCATTA 708
|||||
241 TGCTACCACTGTAAAGTGAAGGGGAAACTGA-TGCCAATCTTTATCTCCAGAAACCATTA 299
|||||
709 GCAGATGGTGTCAAAATCCAGCAATGGCTTGAATACCTGCGCTTATTAAGAGACCTTGG 768
|||||
300 GCAGATGGTGTCAAAATCCAGCAATGGCTTGAACACCTGCGCTTATTAAGAGACCTTGG 359
|||||
769 GATGATGCTTCACTGTCAACCAAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGG 828
|||||
360 GATGATATCTTCACTGTCAACAGAGATGAATTAAGTGTGCAACCCAGCTGGTGTGGGAG 419
|||||
829 AGAATGAACACTGCTCATTCAGCCGACTGCTGGGTGGGCACTGGCTGCTGCTGCTCAG 888
|||||
420 AGGATGAACACTGCTCATTCAGCTGTGGAGTGGTGTGGAGTGGCTGCTGCTGCTCAA 479
|||||
889 CATTTCCTCAACAGTCTCTCCAGAAGTAAAGACGCTGTCATCTGTACTCAGTGGGGGAAT 948
|||||
480 CATTTCCTCAACAGTCTCTCCANAGTAAAGACATTTGTTGCTGCTAGTGGTGAAT 539
|||||
949 GTAGACCTAA---CCTCCCTGAACCTGGGTGGGCGAGGCTGAACGCGACCTCTTACCAG 1005
|||||
540 GTAGACTTAACTCTCTCCATACTTGGGTGAAGCAGGCTGAAGGCGACCTCTTATCAG 599
|||||
1006 ACGGTCTGTTT 1016
|||||
600 TCTGTTTCTGT 610
|||||

RESULT 14
AI322578
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AI322578 455 bp mRNA linear EST 23-DEC-1998
mi51e12.y1 Soares mouse embryo NbMe13.5 14.5 Mus musculus cDNA
clone IMAGE:467086 5' similar to SW:YKVB8.YEAST P36007 HYPOTHETICAL
34.9 KD PROTEIN IN PRE2-JEN1 INTERGENIC REGION. ;, mRNA sequence.
AI322578
AI322578.1 GI:4057007
EST.
house musculus
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 455)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Thelising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:280902
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the

```

```
correct orientation)
Seq primer: -40RP from Gibco
High quality sequence stop: 454.
Location/Qualifiers
I. .455
/organism="Mus musculus"
/db_stain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="IMAGE:467086"
/clone_lib="Soares mouse embryo NBMEL3.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCAATCTGAAGTGGAGCGCGCGAAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT 140 a 121 c 90 g 104 t
ORIGIN
Query Match 44.7%; Score 455; DB 9; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.4e-126;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 18 CATCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCAGACTCTATGCACCTCAC 77
Db 1 CATCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCAGACTCTATGCACCTCAC 60
QY 78 CCAGTGTCTAACAAGCTCCATTTTGAATCAAAATAGCAGGCGCAATCTTTTCTCAATG 137
Db 61 CCAGTGTCTAACAAGCTCCATTTTGAATCAAAATAGCAGGCGCAATCTTTTCTCAATG 120
QY 138 TGAGCTCTTCCAGAAACTGGGTCTTTTAAAGATTCGAGTGCGCCCTTAATGCCATCAGAG 197
Db 121 TGAGCTCTTCCAGAAACTGGGTCTTTTAAAGATTCGAGTGCGCCCTTAATGCCATCAGAG 180
QY 198 CTTAATTCCTGCACGCCAGAGAGAGCCAAAGCCGTAGTTACTCAGCAGCGGAAA 257
Db 181 CTTAATTCCTGCACGCCAGAGAGAGCCAAAGCCGTAGTTACTCAGCAGCGGAAA 240
QY 258 CCATGGCCAAAGCTCTCACCTATGCTGCTAAACTGGAAGGAATTCCTGCTTACATTTGGT 317
Db 241 CCATGGCCAAAGCTCTCACCTATGCTGCTAAACTGGAAGGAATTCCTGCTTACATTTGGT 300
QY 318 TCCCAAAACAGCTCCCAACTGCAAGAACTGGCAATCCAAAGCCTATGGAGCATCGATAGT 377
Db 301 TCCCAAAACAGCTCCCAACTGCAAGAACTGGCAATCCAAAGCCTATGGAGCATCGATAGT 360
QY 378 ATACTGTGACCAAGTGCAGCTCCAGAGAAAGCTCACTCAAGAAATTTATCCAAACAA 437
Db 361 ATACTGTGACCAAGTGCAGCTCCAGAGAAAGCTCACTCAAGAAATTTATCCAAACAA 420
QY 438 AGAAGGCATCTTGGTCCATCCCAACACAGAGCGCTG 472
Db 421 AGAAGGCATCTTGGTCCATCCCAACACAGAGCGCTG 455
RESULT 15
BG965678 729 bp mRNA linear EST 12-JUN-2001
LOCUS 602830522F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4985237 5',
DEFINITION mRNA sequence.
ACCESSION BG965678
VERSION BG965678.1 GI:14353315
```

```
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://imgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue procurement: Jeffrey E. Green, M.D.
CDNA Library preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10992 row: 1 column: 06
High quality sequence stop: 719.
Location/Qualifiers
I. .729
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="IMAGE:4985237"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 211 a 163 c 174 g 181 t
ORIGIN
Query Match 43.8%; Score 445.6; DB 13; Length 729;
Best Local Similarity 98.9%; Pred. No. 1.4e-123;
Matches 459; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 553 GGAGGAGGAGGAATGGTCTGCGATAGCCATTACAAATTAAGCCCTGAACCTAGTGTG 612
Db 1 GGAGGAGGAGGAATGGTCTGCGATAGCCATTACAAATTAAGCCCTGAACCTAGTGTG 60
QY 613 AAGGTATAGCTGCTGAGCCCTCGAATGCAGATGACTGTACAGTCTAACTGAAGGA 672
Db 61 AAGGTATAGCTGCTGAGCCCTCGAATGCAGATGACTGTACAGTCTAACTGAAGGA 120
QY 673 GAAC TGACCCCAACTTTCATCTCCAGAAACCATAGCAGATGGTGTCAAATCCAGCAT 732
Db 121 GAAC TGACCCCAACTTTCATCTCCAGAAACCATAGCAGATGGTGTCAAATCCAGCAT 180
QY 733 GGCTTGAATACCTGGCCTATTATAAGAGACCTTGGATGATGTCTTCACTGTCCAGGA 792
Db 181 GGCTTGAATACCTGGCCTATTATAAGAGACCTTGGATGATGTCTTCACTGTCCAGGA 240
QY 793 GATGAATCAAGTATGCARACCCAGCTGGTGTGGGGAGAAATGAACTGCTCAATGAGCG 852
Db 241 GATGAATCAAGTATGCARACCCAGCTGGTGTGGGGAGAAATGAACTGCTCAATGAGCG 300
QY 853 ACTGCTGGCGTGGCACTGGCTGCAGTGTCTTCAGCAATTTCCAAACAGTCTCTCCAGAA 912
Db 301 ACTGCTGGCGTGGCACTGGCTGCAGTGTCTTCAGCAATTTCCAAACAGTCTCTCCAGAA 360
QY 913 GTAAGAACGCTGCTGCTACTAGTGGGGGAATGTAGACCTTAACCTCCCTGAACTGG 972
Db 361 GTAAGAACGCTGCTGCTACTAGTGGGGGAATGTAGACCTTAACCTCCCTGAACTGG 420
QY 973 GTGGGGCAGGCTGAACGGCCAGCTCTTACCAGAGCGGTCTGTTT 1016
Db 421 GTGGGGCAGGCTGAACGGC-CAGCTCTTACCAGAGCGGTCTGTTT 463
```

```
Search completed: June 24, 2003, 05:35:27
```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:36:57 ; Search time 19.4713 Seconds  
(without alignments)  
722.114 Million cell updates/sec

Title: US-09-889-609B-8  
Perfect score: 1740  
Sequence: 1 MCAQYCISFADVEKAHINIQ.....TSLNWMVQAEAPAPYQTVSV 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1740	100.0	339	1 SRR_MOUSE	O9qzx7 mus musculus
2	1582.5	90.9	340	1 SRR_HUMAN	Q9gzta4 homo sapien
3	587	33.7	326	1 YKNE_YEAST	P36007 saccharomyc
4	560	32.2	323	1 YKNE_SCHPO	O59791 schizosacch
5	491	28.2	329	1 THD2_SALTY	PI1954 salmonella
6	476	27.4	329	1 THD2_ECOLI	P05792 escherichia
7	436	25.1	332	1 Y4TJ_RHISN	P55664 rhizobium s
8	421	24.2	513	1 THD1_PASMU	Q9ckj2 pasteurella
9	405	23.3	513	1 THD1_HAEIN	P46493 haemophilus
10	395.5	22.7	576	1 THDH_YEAST	P00927 saccharomyc
11	384	22.1	514	1 THD1_SALTY	P20506 salmonella
12	379.5	21.8	550	1 THDH_ARKAD	O42615 arxula aden
13	378	21.7	514	1 THD1_ECOLI	P04968 escherichia
14	371.5	21.4	595	1 THD1_LYCES	P25306 lycopersico
15	371	21.3	592	1 THD1_ARKAD	Q9zss6 arabidopsis
16	364	20.9	422	1 THD1_BACSU	P37946 bacillus su
17	362.5	20.8	416	1 THD1_LACLA	Q02145 lactococcus
18	358	20.6	590	1 THD1_CICAR	O39469 cicier ariet
19	352	20.2	507	1 THD1_BURCE	P53607 burkholderi
20	339	19.5	415	1 THD1_BACHD	Q9kx83 bacillus ha
21	323.5	18.6	429	1 THD1_MYCTU	O10766 mycobacteri
22	307	17.6	436	1 THD1_CORGL	Q04513 corynebacte
23	300.5	17.3	427	1 THD1_MYCLE	Q9x7f1 mycobacteri
24	283.5	16.3	362	1 SDHL_RAT	P09367 rattus norv
25	252.5	14.5	328	1 SDHL_HUMAN	P20132 homo sapien
26	218.5	12.6	359	1 THD1_SOLUT	P31212 solanum tub
27	212	12.2	352	1 THRC_BACSP	P09123 bacillus sp
28	208.5	12.0	352	1 THRC_BACSU	P04990 bacillus su
29	208	12.0	312	1 CYSK_SYNY3	P73410 synecocyst
30	202.5	11.6	311	1 CYSK_BACSU	O34476 bacillus su
31	201	11.6	307	1 CYSK_BACSU	P37887 bacillus su
32	200	11.5	338	1 SDHL_YEAST	P17324 saccharomyc
33	198.5	11.4	310	1 CYSK_MYCLE	O32978 mycobacteri

34	198.5	11.4	310	1	CYSK_MYCTU	P95230 mycobacteri
35	198.5	11.4	382	1	THRC_SYNY3	P74193 synecocyst
36	189.5	10.9	307	1	CYSK_FLASP	Q59447 flavobacter
37	187.5	10.8	354	1	THRC_BACHD	O9k7e3 bacillus ha
38	185.5	10.7	550	1	CBS_HUMAN	P35520 homo sapien
39	181	10.4	322	1	CYSK_ARKAD	P47998 arabidopsis
40	177	10.2	322	1	CYK1_BRAJU	O32733 brassica ju
41	177	10.2	325	1	CYSK_CITLA	O43317 citrullus l
42	176	10.1	299	1	CYSM_CAMJE	P71128 campylobact
43	176	10.1	560	1	CBS_RAT	P32232 rattus norv
44	172	9.9	324	1	CYK2_BRAJU	O32735 brassica ju
45	172	9.9	325	1	CYK2_ORVSA	Q9xea8 oryza sativ

ALIGNMENTS

RESULT 1

ID	SRR_MOUSE	STANDARD;	PRT;	339 AA.
AC	O9QZX7;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Serine racemase (EC 5.1.1.-).			
GN	SRR.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/c;			
RX	MEDLINE=20027561; PubMed=10557334;			
RA	Wolosker H., Blackshaw S., Snyder S.H.;			
RT	"Serine racemase: a glial enzyme synthesizing D-serine to regulate			
RT	glutamate-N-methyl-D-aspartate neurotransmission.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:13409-13414(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RA	Strausberg R.;			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: Catalyzes the synthesis of D-serine from L-serine.			
CC	-!- TISSUE SPECIFICITY: Brain.			
CC	-!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.			

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR	EMBL; AF148321; AAF08701.1; .			
DR	EMBL; BC011164; AAH11164.1; .			
DR	HSP; P04968; ITDJ			
DR	MGD; MGI:1351636; Srr.			
DR	InterPro; IPR001926; B6_enzyme_beta.			
DR	InterPro; IPR000634; S/T_dehydrtse.			
DR	Pfam; PF00291; PALP; 1.			
DR	PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.			
KW	Isomerase; Pyridoxal phosphate.			
FT	BINDING 56 PYRIDOXAL PHOSPHATE (BY SIMILARITY).			
SQ	SEQUENCE 339 AA; 36359 MW; B9AE9A933658728 CRC64;			

Query Match 100.0%; Score 1740; DB 1; Length 339;  
Best Local Similarity 100.0%; Pred. No. 6.5e-126;  
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCAQYCISFADVEKAHINIQSDIHLTPVLTSILNQIAGRNLFKCELFQKTSFKIRGA 60  
|||||

Db 1 MCAQYCISFADVEKAHINIQDSIHLPVLTSSILNQIAGRNLPFKCFLFQKTSFKIRGA 60  
 QY 61 LNAIRGLIPDTPPEKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPTAPNCKKLAIQA 120  
 Db 61 LNAIRGLIPDTPPEKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPTAPNCKKLAIQA 120  
 QY 121 YGASIVYCDPSDESREKVTQIMQETGILVHPNQPAPVIAQOGTIALEVLNQVPLVDAL 180  
 Db 121 YGASIVYCDPSDESREKVTQIMQETGILVHPNQPAPVIAQOGTIALEVLNQVPLVDAL 180  
 QY 181 VVPVGGGVMAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGV 240  
 Db 181 VVPVGGGVMAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGV 240  
 QY 241 KSSIGLNTWPIIRDLDVDFVTVEDEIKYATQLVWGRMKLLIEPTAGVLAALVLSQHFQT 300  
 Db 241 KSSIGLNTWPIIRDLDVDFVTVEDEIKYATQLVWGRMKLLIEPTAGVLAALVLSQHFQT 300  
 QY 301 VSPVKNCIVLSGNVDLTSLNWVGQAERPAPYQTVSV 339  
 Db 301 VSPVKNCIVLSGNVDLTSLNWVGQAERPAPYQTVSV 339

## RESULT 2

SRR\_HUMAN  
 ID SRR\_HUMAN STANDARD; PRT; 340 AA.  
 AC Q9GZT4;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Serine racemase (EC 5.1.1.-).  
 GN SRR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20510003; PubMed=11054547;  
 RA Demiranda J., Santoro A., Engelender S., Wolosker H.;  
 RT "Human serine racemase: molecular cloning, genomic organization and  
 functional expression";  
 RL Gene 256:183-188(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Wakatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murekawa K., Ono Y., Takiguchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,  
 RA Ninomiya K., Iwayanagi T.;  
 RT "NEDO human cDNA sequencing project";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Catalyzes the synthesis of D-serine from L-serine.  
 CC -1- TISSUE SPECIFICITY: Brain. Expressed at high levels in hippocampus  
 CC and corpus callosum, intermediate levels in substantia nigra and  
 CC caudate, and low levels in amygdala, thalamus, and subthalamic  
 CC nucleus.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF169974; AAC27081.1;  
 CC EMBL: AK023169; BAB14442.1;  
 CC Genbank: HGNC:14398; SRR.

DR MIM; 606477; -  
 DR InterPro; IPR001926; B6\_enzyme\_beta.  
 DR InterPro; IPR000634; S/T\_dehydrtse.  
 DR Pfam; PF00291; PALP; 1.  
 DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.  
 KW isomerase; Pyridoxal phosphate.  
 FT BINDING 56 56 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 SQ SEQUENCE 340 AA; 36566 MW; 873342C62D5D7B9D CRC64;  
 Query Match 90.9%; Score 1582.5; DB 1; Length 340;  
 Best Local Similarity 89.7%; Pred. No. 7.2e-114;  
 Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;  
 QY 1 MCAQYCISFADVEKAHINIQDSIHLPVLTSSILNQIAGRNLPFKCFLFQKTSFKIRGA 60  
 Db 1 MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQIAGRNLPFKCFLFQKTSFKIRGA 60  
 QY 61 LNAIRGLIPDTPPEKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPTAPNCKKLAIQA 120  
 Db 61 LNAIRGLIPDTPPEKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPTAPNCKKLAIQA 120  
 QY 121 YGASIVYCDPSDESREKVTQIMQETGILVHPNQPAPVIAQOGTIALEVLNQVPLVDAL 180  
 Db 121 YGASIVYCDPSDESREKVTQIMQETGILVHPNQPAPVIAQOGTIALEVLNQVPLVDAL 180  
 QY 181 VVPVGGGVMAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGV 240  
 Db 181 VVPVGGGVMAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGV 240  
 QY 241 KSSIGLNTWPIIRDLDVDFVTVEDEIKYATQLVWGRMKLLIEPTAGVLAALVLSQHFQT 300  
 Db 241 KSSIGLNTWPIIRDLDVDFVTVEDEIKYATQLVWGRMKLLIEPTAGVLAALVLSQHFQT 300  
 QY 301 VSPVKNCIVLSGNVDLTSLNWVGQAERPAPYQTVSV 339  
 Db 301 VSPVKNCIVLSGNVDLTSLNWVGQAERPAPYQTVSV 339

## RESULT 3

YKVB\_YEAST  
 ID YKVB\_YEAST STANDARD; PRT; 326 AA.  
 AC P36007;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 34.9 kDa protein in COS9-JEN1 intergenic region.  
 GN YKL218C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OC NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=95028164; PubMed=7941750;  
 RA Tzeremia M., Horaitis O., Alexandraki D.;  
 RT "The complete sequencing of a 24.6 kb segment of yeast chromosome XI  
 RT identified the known loci URA1, SAC1 and TRP3, and revealed 6 new  
 RT open reading frames including homologues to the threonine  
 RT dehydratases, membrane transporters, hydantoinsases and the  
 RT phospholipase A2-activating protein.";  
 RL Yeast 10:663-679(1994).  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF169974; AAC27081.1;  
 CC EMBL: AK023169; BAB14442.1;  
 CC Genbank: HGNC:14398; SRR.

```

DR EMBL; X75951; CAA53555.1; -.
DR EMBL; Z28218; CAA82063.1; -.
DR PIR; S38061; S38061.
DR PIR; S44320; S44320.
DR HSP; P04968; ITDJ.
DR SGD; S0001701; YKL218C.
DR InterPro; IPR001926; B6_enzyme_beta.
DR InterPro; IPR000634; S/T_dehydrtse.
DR Pfam; PF00291; PALP; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW Hypothetical protein; Lyase; Pyridoxal phosphate. (BY SIMILARITY).
FT BINDING 53 53
SQ SEQUENCE 326 AA; 34899 MW; 21CF7EFC8AB4431 CRC64;

Query Match 33.7%; Score 587; DB 1; Length 326;
Best Local Similarity 40.9%; Pred. No. 8.e-38;
Matches 128; Conservative 66; Mismatches 111; Indels 8; Gaps 4;

QY 8 SFADVEKAHINIODSIHLTPVLTSSILNQIAGRNLFKCELFOKTSFKIRGALNAIRGL 67
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
Db 5 TYGDVLDSNRKEYNKTPTVLTSMNLDRGAIIVKGFNFORVGAFFRGAMNAVSKL 64

QY 68 IPTPEKPKAVYTHSSGNHGQALTYAAKLEGIPAVIVPOTAPNCKKLAIOAYGASIVY 127
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
Db 65 --SDEKRSKGVITAFSSGNHQAIALSAKLLNVPATIVMPEDAPALKVAATAGYGAHIR 121

QY 128 CDPDESREKVTQRIQMOETEGILVHPNQEPVAVIAGOGTIALEVLNOVPLVDALVWPVGG 187
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
Db 122 YNYTEDREQIQRLAAEHCFALIPYDHPDVIAGOGTSKELLEEVGQDLALFVPLGG 181

QY 188 GWVAGIATITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVKSS-IGL 246
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
Db 182 GLLSGSALAARSLSPCKIFGVEPEAGNDGQSQFRSGSIV-HINTPKTIADGAQTQHLGE 240

QY 247 NTWPIIRDVLDDVFTVTEDEIKATQLVGRMKLLIEPTAGVALAAVLSOHFTQVSPYK 306
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
Db 241 YPAIIRNDDILTVSDQELVKCMHFLAERMKVVVEPTACLGAFGALLKKEELVG---K 297

QY 307 NVCIVLSGGNVDL 319
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
Db 298 KVGIIISGGNVDM 310

RESULT 4
YCNE_SCHPO
ID YCNE_SCHPO STANDARD; PRT; 323 AA.
AC Q59791;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical protein C320.14 in chromosome III.
GN SPCC320.14 OR SPCC330.15C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Hamlin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin A., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., O'Neil C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

```

```

RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Sipakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT *The genome sequence of Schizosaccharomyces pombe.*;
RL Nature 415:871-880(2002).
CC -| COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -| SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL022245; CAA18316.1; -.
DR EMBL; AL031603; CAA20920.1; -.
DR HSP; P04968; ITDJ.
DR InterPro; IPR001926; B6_enzyme_beta.
DR InterPro; IPR000634; S/T_dehydrtse.
DR Pfam; PF00291; PALP; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; FALSE_NEG.
KW Hypothetical protein; Lyase; Pyridoxal phosphate. (BY SIMILARITY).
FT BINDING 57 57
SQ SEQUENCE 323 AA; 35048 MW; 21187ESA69FA5348 CRC64;

Query Match 32.2%; Score 560; DB 1; Length 323;
Best Local Similarity 39.6%; Pred. No. 1e-35;
Matches 125; Conservative 68; Mismatches 109; Indels 14; Gaps 7;

QY 8 SFADVEKAHINIODSIHLTPVLTSSILNQIAGRNLFKCELFOKTSFKIRGALNAIRGL 67
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
Db 9 TYDDVASASERIKKFANKPTVLTSSVNFKEFVAEVEFFKCNFMGAFKFRGALNALSQ 68

QY 68 IPTPEKPKAVYTHSSGNHGQALTYAAKLEGIPAVIVPOTAPNCKKLAIOAYGASIVY 127
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
Db 69 ---NEAQRKAGVLTSSGNHQAIALSAKILGIPAKLIIMPLDPAEAKVAATKYGQGVIM 125

QY 128 CDPDESREKVTQRIQMOETEGI-LVHPNQEPVAVIAGOGTIALEVLNOVPLVDALVWPVGG 186
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
Db 126 YDRYKDDREKMAKEI-SEREGLTIPPDHPVLAGQGTAAKELFEVGLDALFVCLGG 184

QY 187 GHWAGIATITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVKSS-IG 245
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
Db 185 GLLSGSALAARHFAFNCEVYGEPEAGNDGQSQFRKGSIV-HIDTPKTIADGAQTQHLG 243

QY 246 LNTWPIIRDVLDDVFTVTEDEIKATQLVGRMKLLIEPTAGVALAAVLSOHFTQVSPYK 305
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
Db 244 NYTFSIIKEKVDILTVSDEELIDCLKFYAARMKIVVEPTGCLSFAAA-----RAMKEKL 298

QY 306 KN--VCIVLSGGNVDL 319
   : : | | | | : : | | | | : : | | | | : : | | | | : : | | | | :
Db 299 KNKRIGIISGGNVDI 314

RESULT 5
THD2_SALTY
ID THD2_SALTY STANDARD; PRT; 329 AA.
AC P11954;
DT 01-OCT-1989 (Rel. 12, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Threonine dehydratase catabolic (EC 4.2.1.16) (Threonine deaminase).

```

GN TDCB OR STM3244.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 LT2.";  
 RL Nature 413:852-856(2001).  
 RN [2]  
 RP SEQUENCE OF 1-24.  
 RX MEDLINE=83023208; PubMed=6751404;  
 RA Kim S.S., Datta P.;  
 RT "Chemical characterization of biodegradative threonine dehydratases  
 from two enteric bacteria";  
 RL Biochim. Biophys. Acta 706:27-35(1982).  
 CC -|- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +  
 H(2)O.  
 CC -|- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -|- ENZYME REGULATION: EACH PROTEIN MOLECULE CAN BIND UP TO FOUR  
 CC MOLECULES OF AMP, WHICH ACTS AS AN ALLOSTERIC ACTIVATOR TO THE  
 CC ENZYME. THE ENZYME IS ALSO INHIBITED BY ALPHA-KETO ACIDS AND OTHER  
 CC CATABOLITES. IT IS ALLOSTERICALLY INHIBITED BY ISOLEUCINE  
 CC AND ALLOSTERICALLY ACTIVATED BY VALINE.  
 CC -|- PATHWAY: Threonine catabolism.  
 CC -|- SUBUNIT: HOMOTETRAMER.  
 CC -|- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AE008849; AAL22117.1;  
 DR PIR; B22317; B22317.  
 DR StGene; SG10390; tdcB.  
 DR InterPro; IPR001926; B6\_enzyme\_beta.  
 DR InterPro; IPR00634; S/T\_dehydrtse.  
 DR Pfam; PF00291; PALP; 1.  
 DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.  
 DR Lyase; Pyridoxal phosphate; Allosteric enzyme; Complete proteome.  
 KW BINDING 58 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 FT CONFLICT 24 24 K -> I (IN REF. 2).  
 FT SEQUENCE 329 AA; 35141 MW; 1C619B021DE817C CRC64;  
 Query Match  
 Best Local Similarity 36.8%; Score 491; DB 1; Length 329;  
 Matches 121; Conservative 68; Mismatches 116; Indels 24; Gaps 8;  
 QY 7 ISADVEKAHINTQSDTHLTPVLTSSLTNIOAGRNLFKCELTQKTSFKIRGALNAIRG 66  
 Db 9 VAIEDILEAKKRLAGIKYTKTGMPSRNSYFSERCKGEIIFLKFENMORTGSEFKIRGAFNKLSS 68  
 QY 67 LIPDTPPEKPAVVTHSSNGHGAALTYAAKLEGIPAYIVVPOTPAPNCKKLAIGASIV 126  
 Db 69 L---TEAKRKGVVACSGAGNHAGVLSLSCAMLGIDGKVVMPKAPKSKVAATCDYSAEVV 125  
 QY 127 -YCDPSDESKVQRIMQETEG-ILVHPNQEPVIAQGQTIALEVLNQVPLVDALVVPV 184  
 Db 126 LHGDNFNDITAKVSEIV--ETEGRIFFTPYDDPKPIAGQGTIGLEIMEDLYDNDVIVPI 193  
 QY 185 GGGGMVAGIAITKALKPSKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVK-SS 243

## RESULT 6

ID THD2\_ECOLI  
 THD2\_ECOLI STANDARD; PRT; 329 AA.  
 AC P05792;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine dehydratase catabolic (EC 4.2.1.16) (Threonine deaminase).  
 GN TDCB OR B3117 OR Z4469 OR ECS3997.  
 OS Escherichia coli, and  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=87092415; PubMed=3540965;  
 RA Datta P., Goss T.J., Ommaas J.R., Patil R.V.;  
 RT "Covalent structure of biodegradative threonine dehydratase of  
 RT Escherichia coli: homology with other dehydratases";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:393-397(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=89282418; PubMed=2660107;  
 RA Schweizer H., Datta P.;  
 RT "The complete nucleotide sequence of the tdc region of Escherichia  
 RT coli.";  
 RL Nucleic Acids Res. 17:3994-3994(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Murata T., Tanaka M.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Nakagawa C., Ogasawara N., Yasunaga T.,  
 RA Iida T., Takami H., Honda T., Sasaki K., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli



```

QY 241 KSSTGL-NTW--PIIRDLDVDDVFTTEDEIKYATQLWGRMKLLIEPTAGVALAALVLSOH 297
DB 237 GGGIGMANWTFQMRALLDDVLVNGEITAAGIRHAYEHQRILEGAGAVGIAULLS-- 294
QY 298 FQTVSPVKNVCIVLSGQNDVL 319
DB 295 -GKYAARGSGVGVLGQNDIM 315

RESULT 8
THDI_PASMU STANDARD; PRT; 513 AA.
AC Q9CKJ2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine
deaminase).
GN ILVA OR PM1624.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella
OC Pasteurella
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from
threonine in a two-step reaction. The first step is a dehydration
of threonine, followed by rehydration and liberation of ammonia.
CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +
H(2)O.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- PATHWAY: Isoleucine biosynthesis; first step.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE006199; AAK03708.1; -
CC HSSP; P04968; ITDJ.
CC InterPro; IPR001926; B6_enzyme_beta.
CC InterPro; IPR000634; S/T_dehydratse.
CC InterPro; IPR001721; ThrDb_C.
CC Pfam; PF00291; PALP; 1.
CC Pfam; PF00585; Thr_dehydrat_C; 2.
CC TIGRfams; TIGR01124; ilva_2Cterm; 1.
CC PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
CC Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;
KW Complete proteome.
FT BINDING 61
SQ SEQUENCE 513 AA; 56288 MW; 3E427ADC54E5FC7C CRC64;

Query Match
Best Local Similarity 34.1%; Pred. No. 7.7e-25;
Matches 106; Conservative 63; Mismatches 134; Indels 8; Gaps 6;

QY 19 IQDSIHLPVLTSSILNQIAGRLNFKELFQKTSFKIRGALNATIRGLPOTPEPKPA 78
DB 24 VYDVAQVTPQDMAKLSRLGNKVFTRKEDRQPVHFSFKLRGAYAMIAGL---SAEQKASG 80
QY 79 VVTHSSGNHQAALTYAAKLEGIPAYIVVPQAPNCKKLAIQAYGASIVYCDPS-DESRK 137
DB 81 VTAASAGNAQGVALSAAKHLGURALIVMPONTPSIKVDVAVRGVGGVLLHGAFNDFEAKAK 140

```

```

QY 138 VTORIMQETGILVHPNQEPAVTAGOGTIALEVLNOVPLVDALVVPVGGGSMVAGIAITI 197
DB 141 AIE-LAESKNMTFIPDFHPAVTAGOGSTAMELLQONSQIDRIFVPVGGGLAAGIAVLI 199
QY 198 KALKPSVKVYAARPSNADDCYQSKLKGELTPNLPHPETIADGVK-SSIGLNTWPIIRDLV 256
DB 200 KQLMPEIKVIGVE-SKDSACLRYALKRAGKPIDLDRVGLFADGVAVVRKIGDETFRVCCQYI 258
QY 257 DQVFTVTEDEIKYATQLWGRMKLLIEPTAGVALAALVLSOHFOTVSPVKNVCIVLSGQ 316
DB 259 DDVVLVDGDEICAAVKDIFENWRAIAEPSGALSLAG-LKKYVKEHNTQGETLVNLSGAN 317
QY 317 VDLTSLNWNVQ 327
DB 318 LNFHTLRYVSE 328

RESULT 9
THDI_HAEIN STANDARD; PRT; 513 AA.
AC P46493;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine
deaminase).
GN ILVA OR HI0738.1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
McKee A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
McKee A.R., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
Utterback T.R., Hanna M.C., Nguyen B.T., Saudek D.M., Brandon R.C.,
Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION.
RA Koonin E.V., Rudd K.E.;
RL Submitted (SEP-1995) to the SWISS-PROT data bank.
CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from
threonine in a two-step reaction. The first step is a dehydration
of threonine, followed by rehydration and liberation of ammonia.
CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +
H(2)O.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- PATHWAY: Isoleucine biosynthesis; first step.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32757; AAC22398.1; -
CC HSSP; P04968; ITDJ.
CC TIGR; HI0738.1; -

```

```

DR InterPro: IPR001926; B6_enzyme_beta.
DR InterPro: IPR000634; S/T_dehydrtse.
DR InterPro: IPR001721; ThrDh_C.
DR Pfam: PF00291; PALP; 1.
DR TIGRfams: TIGR01124; Thr_dehydrat_C; 2.
DR TIGRfams: TIGR01124; ilvA_2Cterm; 1.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
KW Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;
FT Complete proteome.
FT BINDING 63
SQ SEQUENCE 513 AA; 56662 MW; DF42CA8B6FDE4CD7 CRC64;

Query Match 23.3%; Score 405; DB 1; Length 513;
Best Local Similarity 32.6%; Pred. No. 1.3e-23;
Matches 104; Conservative 68; Mismatches 137; Indels 10; Gaps 8;

QY 12 VEKAHINIQDSIHLTPVLTSIIINQIAGRNLFKCELFQKTSFKIRGALNATRGILPDT 71
DB 19 IVKLSRVYEAQVTPLOKMGKLSERLHNNIWKREDQRPVNFSEKLGAYAMISL---$ 75
QY 72 PEEKRAVTHSSGNHGQALTYAAKLEGIPAYIVPQTAPNCKKLAIQAYGASIVYCDPS 131
DB 76 AEQKAAGVIAASAGNHAQGVAAKQGLKALIVMPQNTPSIKVDVAVRGFGGEVLLHGAN 135
QY 132 -DSREKVTQRIQETEGILVHNPQEPVAVTAGOGTTALEVLNQLVDALVVPVGGGMV 190
DB 136 FDEAKAKAIE-LSKEKMTTIPFDHPLVITAGOGTTALEMQLQVADLDYVYVQVGGGLA 194
QY 191 AGIATITKALKPSVKVYAAEPSNADDCYQSKL-KGELTPNLHPPTIADGVK-SSICLNT 248
DB 195 AGVAILLKQWPEIKIIGVE-SKDSACLKAALDKGEPDTH-IGLPADGVAVKRGDET 252
QY 249 WPIRLDLDVFTVTEDEIKYATQLVWGRMKLLIETAGVALAALVLSQHFQTVSPVKVY 308
DB 253 FRLCQVLDMDLVDSDEVCAAMKDLFENVRAEPPSGALGLAG-LKKYVKQNHIEGKNM 311
QY 309 CIVLSGGNVDLTSLNNVGO 327
DB 312 AAILSGANLNFHLRYVSE 330

RESULT 10
THDH_YEAST STANDARD; PRT; 576 AA.
AC P00927;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Threonine dehydratase, mitochondrial precursor (EC 4.2.1.16)
DE (Threonine deaminase).
GN ILV1 OR YER086W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetaceae;
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Kiehlbrand-Brandt M.C., Holmberg S., Petersen J.G.L.,
RA Nilsson-Tillgren T.;
RA "Nucleotide sequence of the gene for threonine deaminase (ILV1) of
RT Saccharomyces cerevisiae."
RL Carlsberg Res. Commun. 49:567-575(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

```

```

CC -I- CATALYTIC ACTIVITY: L-threonine + H(2)O -> 2-oxobutanoate + NH(3) +
CC H(2)O.
CC -I- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -I- ENZYME REGULATION: ISOLEUCINE ALLOSTERICALLY INHIBITS WHILE
CC VALINE ALLOSTERICALLY ACTIVATES THIS ENZYME.
CC -I- PATHWAY: Isoleucine biosynthesis; first step.
CC -I- SUBUNIT: HOMOTETRAMER.
CC -I- SUBCELLULAR LOCATION: Mitochondrial.
CC -I- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M36383; AAA34705.1; -.
CC DR EMBL: X01466; CAA25696.1; -.
CC DR EMBL: U18839; AAB64641.1; -.
CC DR PIR: A01150; DMBYT.
CC DR HSSP: P04968; ITDJ.
CC DR SGP: S0000888; ILV1.
CC DR InterPro: IPR001926; B6_enzyme_beta.
CC DR InterPro: IPR000634; S/T_dehydrtse.
CC DR InterPro: IPR001721; ThrDh_C.
CC DR Pfam: PF00291; PALP; 1.
CC DR Pfam: PF00585; Thr_dehydrat_C; 2.
CC DR TIGRfams: TIGR01124; ilvA_2Cterm; 1.
CC DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
CC KW Isoleucine biosynthesis; Lyase; Mitochondrion; Pyridoxal phosphate;
CC TRANSIT 1 2 MITOCHONDRION (POTENTIAL).
CC FT CHAIN 7 576 THREONINE DEHYDRATASE.
CC FT BINDING 109 109 PYRIDOXAL PHOSPHATE.
CC FT CONFLICT 259 259 I -> T (IN REF. 1).
CC SQ SEQUENCE 576 AA; 63831 MW; 0801BCBD7EEDDC1F CRC64;

Query Match 22.7%; Score 395.5; DB 1; Length 576;
Best Local Similarity 32.5%; Pred. No. 7.9e-23;
Matches 104; Conservative 68; Mismatches 131; Indels 17; Gaps 10;

QY 18 NIQDSIHLTPVLTSIIINQIAGRNLFKCELFQKTSFKIRGALNATRGILPOTPEKPK 77
DB 71 SVYDVINESPISOGVGLSSRLNTNVILKREDLLPVFSFKLRGAYNMIAKL---DDSORNQ 127
QY 78 AVYTHSSGNHGQALTYAAKLEGIPAYIVPQTAPNCKKLAIQAYGASIV-YCDPSDESRE 136
DB 128 GVIACSAGNHAQGVAAKQGLKALIVMPQNTPSIKYQNVSRGLSQVLYGNDGDEAKA 187
QY 137 KVTQRIQETEGIL-VHPNQEPVAVTAGOGTTALEVLNNOVPL---VDALVVPVGGGMVAG 192
DB 188 ECAK---LAERGLTNIPFPDHPYVITAGOGTVAMEILURVETANKKIFAVFPVGGGLIAG 245
QY 193 IAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVK-SSICLNTWPI 251
DB 246 IGAYLKRVPAPHIKIIGVETVDAATLHNSLORNORTP-LPVVGTFADGTSVSMICEETFRV 304
QY 252 IRLVDVDTFTVTEDEIKYATQLVWGRMKLLIETAGVALAALVLSQHFQTVSPVEV---KNV 308
DB 305 AQGVVDEVLVNTDEICAAVKDFEDTRISVPSGALSVAG-MKKYISTVHPEDHTKNT 363
QY 309 CI-VLSSGNVDLTSLNNVGO 327
DB 364 YVPILSGANNFDRLRVSE 383

RESULT 11
THDH_SALTY STANDARD; PRT; 514 AA.
ID THDH_SALTY
AC P20506; O9L6S8;
DT 01-FEB-1991 (Rel. 17, Created)

```

15-JUN-2002 (Rel. 41, Last sequence update)  
 15-JUN-2002 (Rel. 41, Last annotation update)  
 Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine  
 deaminase).  
 GN ILVA OR STM3905 OR STMD1.87.  
 OS Salmonella typhimurium.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88255870; PubMed=3290055;  
 RA Tallon B.E., Little R., Lawther R.P.;  
 RT "Analysis of the functional domains of biosynthetic threonine  
 deaminase by comparison of the amino acid sequences of three  
 wild-type alleles to the amino acid sequence of biodegradative  
 threonine deaminase.";  
 RL Gene 63:245-252(1988).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
 LT2.";  
 RL Nature 413:852-856(2001).  
 [3]  
 RP SEQUENCE OF 1-10 FROM N.A.  
 RX STRAIN=LT2;  
 RA MEDLINE=89326124; PubMed=2473940;  
 RA Lopes J.M., Lawther R.P.;  
 RT "Physical identification of an internal promoter, *ilvAp*, in the  
 RT distal portion of the *ilvGMDA* operon.";  
 RL Gene 76:255-269(1989).  
 CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from  
 CC threonine in a two-step reaction. The first step is a dehydration  
 CC of threonine, followed by rehydration and liberation of ammonia.  
 CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +  
 CC H(2)O.  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -!- ENZYME REGULATION: ISOLEUCINE ALLOSTERICALLY INHIBITS WHEREAS  
 CC VALINE ALLOSTERICALLY ACTIVATES THIS ENZYME.  
 CC -!- PATHWAY: Isoleucine biosynthesis; first step.  
 CC -!- SUBUNIT: HOMOTETRAMER.  
 CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; M26670; AAA27150.1;  
 DR EMBL; AF233324; AAA233479.1;  
 DR EMBL; AE008882; AAL22755.1;  
 DR EMBL; M25498; AAA27151.1;  
 DR PIR; JT0278; DWEBTT.  
 DR HSP; P04968; ITDJ.  
 DR StryGene; SG10179; *ilvA*.  
 DR InterPro; IPR001926; B6\_enzyme\_beta.  
 DR InterPro; IPR000634; S/T\_dehydrtse.  
 DR InterPro; IPR001721; ThrDh\_C.  
 DR Pfam; PF00291; PALP; 1.  
 DR Pfam; PF00585; Thr\_dehydrat\_C; 2.  
 DR TIGRams; TIGR01124; *ilvA\_2*Term; 1.  
 DR PROSITE; PS00185; DEHYDRATASE\_SER\_THR; 1.  
 CC Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;

KW Allosteric enzyme; Complete proteome.  
 FT BINDING 62 PYRIDOXAL PHOSPHATE.  
 FT CONFLICT 71 A -> T (IN REF. 1).  
 FT CONFLICT 124 F -> L (IN REF. 1).  
 FT CONFLICT 339 A -> G (IN REF. 1).  
 FT CONFLICT 342 A -> T (IN REF. 1).  
 FT CONFLICT 351 SFL -> NEP (IN REF. 1).  
 SQ SEQUENCE 514 AA; 56253 MW; 136BC535F1F0035B CRC64;  
  
 Query Match 22.1%; Score 384; DB 1; Length 514;  
 Best Local Similarity 32.3%; Pred. No. 5.2e-22;  
 Matches 102; Conservative 69; Mismatches 127; Indels 18; Gaps 8;  
  
 QY 19 IQDSIHLPVLTSSILLNQIAGRNLFKCELFQKTSKIRGALNIRGLIPDTPPEKPKA 78  
 Db 25 VYEAQVTPLOKMEKJSSRLDNVILVKREDRQPVSEKRLGAYAMMAGL---TEEQKAHG 81  
 QY 79 VVTHSSGNHGQALTYAAKLEGPAYIVVPTAPNCKKLAIQAYGASIVYCDPS-DESREK 137  
 Db 82 VITASAGNHAQGVAFSSARLGVKSLIVMPKATADIKVDVAVRGFGEVLLHGAFDEAKAK 141  
 QY 138 VTQRIWQETEGILVHPNQEPVIAAGCTIALEVLNQVPLVDALVVPVGGGMVAGIAITI 197  
 Db 142 AIE-LAQOQGTWVPFDPHPMVIAGCTLALELLQDQSHLDLDRVFPVVGGLAAGVAALI 200  
 QY 198 KALKPSVKVYAAEPSNADD--CYOSKLGELTPNLHPPTIADGVK-SSIGLNTWPIIRD 254  
 Db 201 KQMPQIKVIAVE---AEDSACLKALEAGHPVDLPVGLFAGGVAVKRVGDETFRCQE 257  
 QY 255 LVDDVFVTDEIKYATQLVWGRMKLLIEPTAGVAAAV---LSQHFQTVSPEVKNVCIV 311  
 Db 258 YLDDIITVDSDAICAAMKDLFEDVRAVAPSGALAGMKKYIAQH---NIRGERLAHV 313  
 QY 312 LSGGNVDLTLNHWGQ 327  
 Db 314 LSGANYNFHGLRYVSE 329  
  
 RESULT 12  
 THDH\_ARXAD  
 ID THDH\_ARXAD STANDARD; PRT; 550 AA.  
 AC 042615;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine dehydratase, mitochondrial precursor (EC 4.2.1.16)  
 DE (Threonine deaminase).  
 GN ILV1.  
 OS Arxula adeninivorans.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; mitosporic Dipodascaceae; Arxula.  
 OX NCBI\_TaxID=37620;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LS3;  
 RX MEDLINE=98398049; PubMed=9730281;  
 RA Wartmann T., Roessel H., Kunze I., Bode R., Kunze G.;  
 RT "ALLV1 gene from the yeast *Arxula adeninivorans* LS3 -- a new selective  
 RT transformation marker.";  
 RL Yeast 14:1017-1025(1998).  
 CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +  
 CC H(2)O.  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -!- PATHWAY: Isoleucine biosynthesis; first step.  
 CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial







**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:36:57 ; Search time 19,5287 Seconds  
(without alignments)  
722.114 Million cell updates/sec

Title: US-09-889-609B-10

Perfect score: 1735

Sequence: 1 MCAQYCI5FADVEKAHINR.....SSITWVQAERPASQSVSV 340

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1735	100.0	340	1 SRR_HUMAN	Q9gzt4 homo sapien
2	1582.5	91.2	339	1 SRR_MOUSE	Q9gzt7 mus musculus
3	605	34.9	326	1 YKVB_YEAST	P36007 saccharomyc
4	565	32.6	323	1 YCNE_SCHPO	O59791 schizosacch
5	471	27.1	329	1 THD2_SALTY	P11954 salmonella
6	457	26.3	329	1 THD2_ECOLI	P05792 escherichia
7	445	25.6	332	1 Y4TJ_RHISN	P55664 rhizobium s
8	404	23.3	513	1 THD1_PASMU	Q9ckj2 pasteurellia
9	395.5	22.8	513	1 THD1_HAEIN	P46493 haemophilus
10	390.5	22.5	576	1 THDH_YEAST	P00927 saccharomyc
11	376	21.7	422	1 THD1_BACSU	P37946 bacillus su
12	375	21.6	507	1 THD1_BURCE	P53607 burkholderi
13	369.5	21.3	550	1 THDH_ARXAD	O42615 arxula aden
14	368	21.2	592	1 THD1_ARATH	Q92856 arabidopsis
15	363	20.9	595	1 THD1_LYCES	P25306 lycopersico
16	362	20.9	514	1 THD1_SALTY	P20506 salmonella
17	357	20.6	514	1 THD1_ECOLI	Q04968 escherichia
18	351.5	20.3	416	1 THD1_LACLA	Q02145 lactococcus
19	351	20.2	590	1 THD1_CICAR	Q93469 cicar ariet
20	347	20.0	415	1 THD1_BACHD	Q9kc63 bacillus ha
21	314.5	18.1	429	1 THD1_MYCTU	Q10766 mycobacteri
22	311.5	18.0	436	1 THD1_CORGL	Q04513 corynebacte
23	301.5	17.4	427	1 THD1_MYCLE	Q9x7f1 mycobacteri
24	267.5	15.4	362	1 SDHL_RAT	P09367 rattus norv
25	265.5	15.3	328	1 SDHL_HUMAN	P20132 homo sapien
26	210	12.1	359	1 THD1_SOLTU	P31212 solanum tub
27	209.5	12.1	311	1 CYSM_BACSU	O34476 bacillus su
28	203	11.7	352	1 THRC_BACSP	P09123 bacillus sp
29	201.5	11.6	354	1 THRC_BACHD	Q9k7e3 bacillus ha
30	201	11.6	307	1 CYSK_BACSU	P37887 bacillus su
31	200.5	11.6	352	1 THRC_BACSU	P04990 bacillus su
32	200	11.5	382	1 CYSK_SYNY3	P74193 synechocyst
33	198.5	11.4	312	1 CYSK_SYNY3	P73410 synechocyst

RESULT 1  
SRR\_HUMAN  
ID SRR\_HUMAN STANDARD; PRT: 340 AA.  
AC Q9GZT4;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Serine racemase (EC 5.1.1.1).  
GN SRR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Brain;  
TX MEDLINE=20510003; PubMed=11054547;  
RA DeMiranda J., Santoro A., Engelender S., Wolosker H.;  
RT "Human serine racemase: molecular cloning, genomic organization and  
functional expression.";  
RL Gene 256:183-188(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
RA Ninomiya K., Iwayanagi T.;  
RL "NEO human cDNA sequencing project.";  
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Catalyzes the synthesis of D-serine from L-serine.  
CC -!- TISSUE SPECIFICITY: Brain. Expressed at high levels in hippocampus  
and corpus callosum, intermediate levels in substantia nigra and  
caudate, and low levels in amygdala, thalamus, and subthalamic  
nuclei.  
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: AF169974; AG27081.1; -  
CC EMBL: AK023169; BAB14442.1; -  
CC Genew; HGNC:14398; SRR.  
CC MIM; 606477; -  
CC InterPro: IPR001926; B6\_enzyme\_beta.  
CC InterPro: IPR000634; S/T\_dehydrtse.  
CC Pfam; PF00291; PALP; 1.  
CC PROSITE; P500165; DEHYDRATASE\_SER\_THR; 1.  
CC Isomerase; Pyridoxal phosphate.

Oct 27, 2000

```
FT BINDING 56 56 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 340 AA; 36566 MW; 873342C62D5D7B9D CRC64;

Query Match 100.0%; Score 1735; DB 1; Length 340;
Best Local Similarity 100.0%; Pred. No. 1e-125;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCAQYICISFADVEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGA 60
DB 1 MCAQYICISFADVEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGA 60
QY 61 LNAVRLVDPALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPPQAPDCCKLAIQA 120
DB 61 LNAVRLVDPALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPPQAPDCCKLAIQA 120
QY 121 YGASIVYCPDSERENAKRVTEETEGIMVHPNQEPAVIAGOGTTALEVLNQVPLVDAL 180
DB 121 YGASIVYCPDSERENAKRVTEETEGIMVHPNQEPAVIAGOGTTALEVLNQVPLVDAL 180
QY 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYOSKLGKMLPNLYPPETIADGV 240
DB 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYOSKLGKMLPNLYPPETIADGV 240
QY 241 KSSIGLNTWPIIRDLVDDIFTVTDEIKCATOLVWERMKLLIEPTAGVGVAAVLSQHFTQ 300
DB 241 KSSIGLNTWPIIRDLVDDIFTVTDEIKCATOLVWERMKLLIEPTAGVGVAAVLSQHFTQ 300
QY 301 VSPEVKNICIVLSGGNVDLTSSITWVKQAEPAASYQSUSV 340
DB 301 VSPEVKNICIVLSGGNVDLTSSITWVKQAEPAASYQSUSV 340

RESULT 2
SRR_MOUSE STANDARD; PRT; 339 AA.
ID SRR_MOUSE
AC Q9QZK7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Serine racemase (EC 5.1.1.-).
GN SRR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=20027561; PubMed=10557334;
RA Wolosker H., Blackshaw S., Snyder S.H.;
RT "Serine racemase: a glial enzyme synthesizing D-serine to regulate
RT glutamate-N-methyl-D-aspartate neurotransmission.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13409-13414(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Catalyzes the synthesis of D-serine from L-serine.
CC -!- TISSUE SPECIFICITY: Brain.
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF148321; AAF08701.1; -
CC EMBL: BC01164; AAH1164.1; -
CC HSSP: P04968; ITDQ.
```

```
DR MGD; MGI:1351636; Srr.
DR InterPro: IPR001926; B6_enzyme_beta.
DR InterPro: IPR000634; S/T_dehydrtse.
DR Pfam: PF00291; PALP; 1.
DR PROSITE; P500165; DEHYDRATASE_SER_THR; 1.
KW Isomerase; Pyridoxal phosphate.
FT BINDING 56 56 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 339 AA; 36359 MW; B9AE9A9336358728 CRC64;

Query Match 91.2%; Score 1582.5; DB 1; Length 339;
Best Local Similarity 89.7%; Pred. No. 4.9e-114;
Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;

QY 1 MCAQYICISFADVEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGA 60
DB 1 MCAQYICISFADVEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGA 60
QY 61 LNAVRLVDPALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPPQAPDCCKLAIQA 120
DB 61 LNAVRLVDPALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPPQAPDCCKLAIQA 120
QY 121 YGASIVYCPDSERENAKRVTEETEGIMVHPNQEPAVIAGOGTTALEVLNQVPLVDAL 180
DB 121 YGASIVYCPDSERENAKRVTEETEGIMVHPNQEPAVIAGOGTTALEVLNQVPLVDAL 180
QY 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYOSKLGKMLPNLYPPETIADGV 240
DB 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYOSKLGKMLPNLYPPETIADGV 240
QY 241 KSSIGLNTWPIIRDLVDDIFTVTDEIKCATOLVWERMKLLIEPTAGVGVAAVLSQHFTQ 300
DB 241 KSSIGLNTWPIIRDLVDDIFTVTDEIKCATOLVWERMKLLIEPTAGVGVAAVLSQHFTQ 300
QY 301 VSPEVKNICIVLSGGNVDLTSSITWVKQAEPAASYQSUSV 340
DB 301 VSPEVKNICIVLSGGNVDLTSSITWVKQAEPAASYQSUSV 339

RESULT 3
YKV8_YEAST STANDARD; PRT; 326 AA.
ID YKV8_YEAST
AC P36007;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 34.9 kDa protein in COS9-JEN1 intergenic region.
GN YKL218C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=95028164; PubMed=7941750;
RA Tzeremia M., Horaitis O., Alexandraki D.;
RT "The complete sequencing of a 24.6 kb segment of yeast chromosome XI
RT identified the known loci URA1, SAC1 and TRP3, and revealed 6 new
RT open reading frames including homologues to the threonine
RT dehydratases, membrane transporters, hydantoinsases and the
RT phospholipase A2-activating protein.";
RL Yeast 10:663-679(1994).
CC -!- COPACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF148321; AAF08701.1; -
CC EMBL: BC01164; AAH1164.1; -
CC HSSP: P04968; ITDQ.
```

DR EMBL: X75951; CAA53555.1; -  
 DR EMBL: 228218; CAA82063.1; -  
 DR PIR: S38061; S38061.  
 DR PIR: S44320; S44320.  
 DR HSP: P04968; 1TDJ.  
 DR SGD: S0001701; YKL218C.  
 DR InterPro: IPR001926; B6\_enzyme\_beta.  
 DR InterPro: IPR000634; S/T\_dehydrtse.  
 DR Pfam: PF00291; PALP; 1.  
 DR PROSITE: PS00165; DEHYDRATASE\_SER\_THR; 1.  
 KW Hypothetical protein; Lyase; Pyridoxal phosphate. (BY SIMILARITY).  
 FT BINDING 53 53  
 SQ SEQUENCE 326 AA; 34899 MW; 21CFEFC9AB4431 CRC64;

Query Match 34.9%; Score 605; DB 1; Length 326;  
 Best Local Similarity 41.0%; Pred. No. 3.2e-39;  
 Matches 133; Conservative 64; Mismatches 119; Indels 8; Gaps 4;

QY 8 SPADVEKAHINTRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRSL 67  
 Db 5 TYGVDLADSNRIKEYVKNTPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRSL 64  
 QY 68 VPDALERPKAVVTHSSNGHQAALTYAAKLEGIPAYIVVPTAPDCKKLAIOAGYASIVY 127  
 Db 65 ----SDEKRSKGVIATFSNGHQAALTYAAKLEGIPAYIVVPTAPDCKKLAIOAGYASIVY 121  
 QY 128 CEPDSRENVAKRVTEETEGIMVHPNOBPAVIAAGGTIALEVNOVPLVDALVVPVGGG 187  
 Db 122 YNRYTEDREIQGLAAEHGFALIPYDHPDVIAGQTSKELLEVEVGQGLDALFVPLGGG 181  
 QY 188 GMLAGIATVTKALKPSVKVYAAEPSNADDCYOSKLGKLMPLNLYPPETIADGVKSS-IGL 246  
 Db 182 GLLSGSALAARSLSPCKIFGVEPEAGNDGQOSFRSGSIV-HINTPKTIADGAQTQHLGE 240  
 QY 247 NTWPIIRDLVDVDFITVTEDEIKATQLVWERKMLLIEPTAGVGVAAVLSQHFQTVSPVK 306  
 Db 241 YFAIIRDNVDLITVSQELVKMHFLAERMKVVEPTACLGFGAGALKKEELVG--K 297  
 QY 307 NTCIVLSGGNVDLTSSITWVQAE 330  
 Db 298 KVGIIILSGGNVDMKRYATLISGKE 321

## RESULT 4

YCNESCHPO  
 ID YCNE SCHPO STANDARD; PRT: 323 AA.  
 AC O59791;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C320.14 in chromosome III.  
 GN SPC320.14 OR SPC330.15C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 ON NCBI\_TaxID=4896;  
 RX STRAIN=972;  
 RC SEQUENCE FROM N.A.  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Williams R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Furnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 DR EMBL: AL022245; CAA18316.1; -  
 DR EMBL: AL031603; CAA20920.1; -  
 DR HSP: P04968; 1TDJ.  
 DR InterPro: IPR001926; B6\_enzyme\_beta.  
 DR InterPro: IPR000634; S/T\_dehydrtse.  
 DR Pfam: PF00291; PALP; 1.  
 DR PROSITE: PS00165; DEHYDRATASE\_SER\_THR; FALSE\_NEG.  
 KW Hypothetical protein; Lyase; Pyridoxal phosphate. (BY SIMILARITY).  
 FT BINDING 57 57  
 SQ SEQUENCE 323 AA; 35048 MW; 21187E9A69FA5348 CRC64;

Query Match 32.6%; Score 565; DB 1; Length 323;  
 Best Local Similarity 40.3%; Pred. No. 3.7e-36;  
 Matches 127; Conservative 67; Mismatches 109; Indels 12; Gaps 6;

QY 8 SPADVEKAHINTRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRSL 67  
 Db 9 TYDDVASASERIKKFKANKTPVLTSSITVKNFEVAEVFKCFKMGAFKRGALNALSQL 68  
 QY 68 VPDALERPKAVVTHSSNGHQAALTYAAKLEGIPAYIVVPTAPDCKKLAIOAGYASIVY 127  
 Db 69 --NEAQRK-AGVLTSSNGHQAALTYAAKLEGIPAYIVVPTAPDCKKLAIOAGYASIVY 125  
 QY 128 CEPDSRENVAKRVTEETEGIMVHPNOBPAVIAAGGTIALEVNOVPLVDALVVPVGGG 187  
 Db 126 YDRYKDDREKMAKEISEREGLTIIPYDHPVLAGGTAAKELFEVGLDALFVCLGGG 185  
 QY 188 GMLAGIATVTKALKPSVKVYAAEPSNADDCYOSKLGKLMPLNLYPPETIADGVKSS-IGL 246  
 Db 186 GLLSGSALAARHFAFPNCEVYGEPEAGNDGQOSFRKGSIV-HIDTPKTIADGAQTQHLGN 244  
 QY 247 NTWPIIRDLVDVDFITVTEDEIKATQLVWERKMLLIEPTAGVGVAAVLSQHFQTVSPVK 306  
 Db 245 YTFSIKKEKVDLITVSDELDLCLFYAARMKIVVEPTGCLSLFAAA-----RAMKEK 299  
 QY 307 N-ICIVLSGGNVDL 319  
 Db 300 NKRIIGIISGGNVDI 314

RESULT 5  
 THD2\_SALTY  
 ID THD2\_SALTY STANDARD; PRT: 329 AA.  
 AC P11954;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine dehydratase catabolic (EC 4.2.1.16) (Threonine deaminase).

GN TDCB OR STM3244.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Letreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RL Nature 413:852-856(2001).  
RN [2]  
RP SEQUENCE OF 1-24;  
RX MEDLINE=83023208; PubMed=6751404;  
RA Kim S.S., Datta P.;  
RT "Chemical characterization of biodegradative threonine dehydratases  
RT from two enteric bacteria.";  
RL Biochim. Biophys. Acta 706:27-35(1982).  
CC -|- CATALYTIC ACTIVITY: L-threonine + H(2)O -> 2-oxobutanoate + NH(3) +  
CC H(2)O.  
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -|- ENZYME REGULATION: EACH PROTEIN MOLECULE CAN BIND UP TO FOUR  
CC MOLECULES OF AMP, WHICH ACTS AS AN ALLOSTERIC ACTIVATOR TO THE  
CC ENZYME. THE ENZYME IS ALSO INHIBITED BY ALPHA-KETO ACIDS AND OTHER  
CC CATABOLITES. IT IS ALLOSTERICALLY INHIBITED BY ISOLEUCINE  
CC AND ALLOSTERICALLY ACTIVATED BY VALINE.  
CC -|- PATHWAY: Threonine catabolism.  
CC -|- SUBUNIT: HOMOTETRAMER.  
CC -|- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AE008849; AAL22117.1; --  
CC PIR: B2317; B2317.  
CC StyGene: SGI0390; tdcB.  
CC InterPro: IPR001926; B6\_enzyme\_beta.  
CC InterPro: IPR000634; S/T\_dehydratse.  
CC Pfam: PF00291; PALP; 1.  
CC PROSITE: PS00165; DEHYDRATASE\_SER\_THR; 1.  
KW Lyase; Pyridoxal phosphate; Allosteric enzyme; Complete proteome.  
FT BINDING 58 58 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
FT CONFLICT 24 24 K -> I (IN REF. 2).  
SQ SEQUENCE 329 AA; 35141 MW; C1C619B021DE817C CRC64;  
Query Match 27.1%; Score 471; DB 1; Length 329;  
Best Local Similarity 36.6%; Pred. No. 5.9e-29;  
Matches 123; Conservative 64; Mismatches 121; Indels 28; Gaps 10;  
QY 7 ISFADVEKAHNIIRDSIHPTVLTSTLSLNQLTGRNLFKCELFQKTSKIRGALNAVR 66  
DB 9 VAIEDILEAKRLAGKIYKTPRMSNYFSERCKGEIFLKFENNQRTGSPKIRGAENKLS 68  
QY 67 LVPDALERPKAVVTHSSNHOALTYAAKLGIPAYIVVQTPADCKKLATQAGASIV 126  
DB 69 LFE---AERKKGWACSGNHAQGVSLSCAMLGIDGKVVMPKGPASKVAATCDYSAEV 125  
QY 127 YCEPDESRENKVRTE--ETEG-IMVHPNPEVAGQGTIALEVLNOVPLVDALVVP 183  
DB 126 L--HGDNFNDTAK-VSEIVETEGRIIPPYDDPKVIAQGGTIGLEIMEDLYDENVIVP 182  
QY 184 VGGGMLAGIATVTKALPKSVKYAAEPSNADDCYOSKLGKMLPNLYPPETIADGVK-S 242

Db 183 IGGGGLIAGIATAIKSINPTIKVIGQAEVNHGMAASYVTGETTH-RTTGTADCCDVS 241  
QY 243 STGLNTWPIIRDIVDDIFTVTEDEIKCATQLVWRMKLLIETAGVGVAAVLS----- 295  
DB 242 RPNLTYYEIVRELVDDIVLVSEDEIRNSMIALIQRNKVITEGAGALACAAALLSGKLDHSI 301  
QY 296 QHFQTVSPVKNTICVLGSGNVDLT--SSITVVKQA 329  
DB 302 QNRKTVS-----IISGNNIDLSRVSOITGLVDA 329  
RESULT 6  
THD2\_ECOLI  
ID THD2\_ECOLI STANDARD; PRT; 329 AA.  
AC P05792;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Threonine dehydratase catabolic (EC 4.2.1.16) (Threonine deaminase).  
GN TDCB OR B3117 OR Z4469 OR ECS3997.  
OS Escherichia coli, and  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562, 83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / W3110;  
RX MEDLINE=87092415; PubMed=3540965;  
RA Datta P., Goss T.J., Omnaas J.R., Patil R.V.;  
RT "Covalent structure of biodegradative threonine dehydratase of  
RT Escherichia coli: homology with other dehydratases.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:393-397(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / W3110;  
RX MEDLINE=89282418; PubMed=2660107;  
RA Schweizer H., Datta P.;  
RT "The complete nucleotide sequence of the tdc region of Escherichia  
RT coli.";  
RL Nucleic Acids Res. 17:3994-3994(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
RL Nature 409:529-533(2001).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli



QY 238 DGKSSIGL-NTW--PIIRDVLVDIETVTEDEIKCATQLVWERMKLLIEPTAGVGAVAL 294  
 Db 234 DSLGGGIGNANWTFQMCRAALLDDVVLNVEGEIAAGIRHAYEHERQILEGAGAVGIAALL 293  
 QY 295 SQHFQTVSEVKNICIVLSGGNDL 319  
 Db 294 S---GRVAARGSGVGVLSGQNDM 315

RESULT 8  
 THDL\_PASMU STANDARD; PRT; 513 AA.  
 AC Q9CKJ2;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine  
 deaminase).  
 GN ILVA OR PM1624.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Pasteurella.  
 OX NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Pm70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from  
 threonine in a two-step reaction. The first step is a dehydration  
 of threonine, followed by rehydration and liberation of ammonia.  
 CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +  
 H(2)O.  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -!- PATHWAY: Isoleucine biosynthesis; first step.  
 CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE006199; AAK03708.1; -  
 CC HSSP; P04968; ITDJ.  
 CC InterPro; IPR001926; B6\_enzyme\_beta.  
 CC InterPro; IPR000634; S/T\_dehydrtse.  
 CC InterPro; IPR001721; ThrDh\_C.  
 CC Pfam; PF00291; PALP; 1.  
 CC Pfam; PF00585; Thr\_dehydrat\_C; 2.  
 CC TIGRFAMs; TIGR01124; ilva\_2cTerm; 1.  
 CC PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.  
 CC Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;  
 CC Complete proteome.  
 FT BINDING 61 PYRIDOXAL PHOSPHATE.  
 SQ SEQUENCE 513 AA; 56288 MW; 3E427ADC54E5FC7C CRC64;

Query Match 23.3%; Score 404; DB 1; Length 513;  
 Best Local Similarity 32.6%; Pred. No. 1.4e-23;  
 Matches 98; Conservative 65; Mismatches 132; Indels 6; Gaps 4;

QY 19 IRDSIHLTPVLSSILNQTLGNRLFKBELFKTGSFKTRGNVAVRSVLPDALEKPKA 78  
 Db 24 VYDVAQVTPLODMAKLSERLGNKVFTRKDRQPVHSFKLRGAYAMTAGL---SAEQKASG 80  
 QY 79 VVTHSSGNHGOALTAAKLEGIPAYIVVPQAPDCCKKLAIOAGSIVYCEPSDESRENV 138  
 Db 81 VTAASAGNRAQGVALSAAKHLGRLALIVMPONTSPISKVDVAVRGFGEVLLHGANFDEAKA 140

QY 139 AKRVTEETEGIMVHPNQEPAVIAOGCTIALEVLNVPLVDALVVPVGGGMLAGIAITVK 198  
 Db 141 ATELAESKNWTFIPPDHPAVIAGQGSIAEMELLOQNSQIDRTFVPVGGGGLAAGIAVLK 200  
 QY 199 ALKPSVYVYAEPNSNADDCYQSKLGLMPLNYPPIETIADGVK-SSIGLNTWPIIRDVLVD 257  
 Db 201 QLMPEIKVIGVESKDSACLYRALKAGKPI-DLDRVGLFADGVAVKRGIDETFRVCCQYID 259  
 QY 258 DIPTVTEDEIKCATQLVWERMKLLIEPTAGVGAVALSOFHQTVSEVKNICIVLSGGNV 317  
 Db 260 DVVLVDGDEICAAVKDIFENRAIEPSPGALSGLAG-LKKYVKEHNITQGETLVNLSGANL 318  
 QY 318 D 318  
 Db 319 N 319

RESULT 9  
 THDL\_HAEIN STANDARD; PRT; 513 AA.  
 AC P46493;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine  
 deaminase).  
 GN ILVA OR HI0738.1.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 RT Science 269:496-512(1995).  
 RN [2]  
 RP IDENTIFICATION.  
 RA Koonin E.V., Rudd K.E.;  
 RL Submitted (SEP-1995) to the SWISS-PROT data bank.  
 CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from  
 threonine in a two-step reaction. The first step is a dehydration  
 of threonine, followed by rehydration and liberation of ammonia.  
 CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +  
 H(2)O.  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -!- PATHWAY: Isoleucine biosynthesis; first step.  
 CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U32757; AAC22398.1; -  
 CC HSSP; P04968; ITDJ.  
 CC TIGR; HI0738.1; -



DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine  
 DE deaminase).  
 GN ILVA.  
 OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OC NCBI\_TaxID=1423;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Arnpriester J.M. Jr., Fink P.S.;  
 RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriess R., Boursier L., Brans A., Braun M., Brignell S., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Gim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*.";  
 RL Nature 390:249-256(1997).  
 CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from  
 CC threonine in a two-step reaction. The first step is a dehydration  
 CC of threonine, followed by rehydration and liberation of ammonia.  
 CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanate + NH(3) +  
 CC H(2)O.  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -!- PATHWAY: Isoleucine biosynthesis; first step.  
 CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; M58606; AAA22549.1; -;  
 DR EMBL; L77246; AAA96639.1; -;  
 DR EMBL; Z99115; CAB14095.1; -;  
 DR HSP; P04968; ITDJ.  
 DR Subtilisin; BG10673; ILVA.  
 DR InterPro: IPR001926; B6\_enzyme\_beta.  
 DR InterPro: IPR000634; S/T\_dehydrtse.  
 DR InterPro: IPR001721; ThrDh\_C.  
 DR Pfam; PF00291; PALP; 1.  
 DR Pfam; PF00585; Thr\_dehydrat\_C; 1.  
 DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.  
 KW Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;  
 KW Complete proteome.  
 FT BINDING 60 60 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 FT CONFLICT 15 15 H -> D (IN REF. 2).  
 FT CONFLICT 139 139 V -> A (IN REF. 2).  
 SQ SEQUENCE 422 AA; 46701 MW; 3027A5ED87084140 CRC64;  
 Query Match 21.7%; Score 376; DB 1; Length 422;  
 Best Local Similarity 31.1%; Pred. No. 15e-21;  
 Matches 99; Conservative 64; Mismatches 143; Indels 12; Gaps 7;  
 QY 7 ISPADVEKAHINIRDSIHILTPVLTSSILNLTGRNLTKCELFCKTGSGFKIRGALNAVR.66  
 Db 11 IQVKHILKAKHONVKQVVIHTPLQRNDRLSERYECNLYLKREDLQVVRFSFKRGAYHKMQ 70  
 QY 67 LVPDALEKRPKAVVTHSSNGHGAALTYAAKLEGIPAYIVVPTQAPDCKKLAIOYGASIV 136  
 Db 71 L--SSQETENGVCVCSAGNHAQGVAFSKHLGIHKIFMPSTTPROKVSQVLEFGKFI 127  
 QY 127 YCPSPDESRENKVRVTE--ETEG-IMVHPNQPAVITAGOGTIALEVLNOVPL-VDAVV 182  
 Db 128 DIILGTGDFDDVYKSAECCAEASRTFIHFDPDVMWAGOGTLAVEILNDIDTEPHLFA 187  
 QY 183 PVGGGMLAGIATVYKALPKSVKYAAEPSNADDCYOSKLGKLMPLNLYPPETIADGVK - 241  
 Db 188 SVGGGGLSGVGYLKNVSPDTKVIAPVAPAGAASYFESNKGHWV-TLOKIDKFDVGA 246  
 QY 242 SSTGLNTWPIIRDVLVDITVTDEIKCATQLVYERMKLLIEPTAGVGVAAVLSDHPTV 301  
 Db 247 KKGTEFTFTLETVDVDDILLVPEGKVTCTSILEYNECAVVAEPAGALSVAALDYKDI - 305  
 QY 302 SPEVKNICIVLSGGNVDL 319  
 Db 306 --KGKNVCVVGSGNNDI 321  
 RESULT 12  
 THD1\_BURCE STANDARD; PRT; 507 AA.  
 AC P53607;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine  
 DE deaminase).  
 GN ILVA.  
 OS *Burkholderia cepacia* (Pseudomonas cepacia).  
 OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
 OC *Burkholderia*.  
 OC NCBI\_TaxID=292;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=17616;  
 RA Bartell J.B., Lessie T.G.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from  
 CC threonine in a two-step reaction. The first step is a dehydration  
 CC of threonine, followed by rehydration and liberation of ammonia.  
 CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanate + NH(3) +  
 CC H(2)O.  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.





Job time : 20.5287 secs

```
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 52-61.
RX STRAIN=CV, TINY, TIM LA154;
RA MEDLINE=91187855; PubMed=2011578;
RT Samach A., Hareven D., Gutfinger T., Ken-Dror S., Lifschitz E.;
   "Biosynthetic theanine deaminase gene of tomato: isolation,
   structure, and upregulation in floral organs.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2678-2682(1991).
CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +
   H(2)O.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- ENZYME REGULATION: ALLOSTERICALLY INHIBITED BY ISOLEUCINE.
CC -!- PATHWAY: Isoleucine biosynthesis; first step.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- MISCELLANEOUS: EXPRESSION IN MATURE FLOWERS IS INCREASED OVER
   500-FOLD.
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M61914; AAA34171.1; -.
DR EMBL: M61915; AAA68097.1; -.
DR PIR: A38628; A38628.
DR HSSP: P04968; 1TDJ.
DR InterPro: IPR001926; B6 enzyme beta.
DR InterPro: IPR000634; S/T dehydratase.
DR InterPro: IPR001721; ThrDh_C.
DR Pfam: PF00291; PALP; 1.
DR Pfam: PF00585; Thr_dehydrat_C; 2.
DR TIGRFAMs: TIGR01124; ilvA_2Cterm; 1.
DR PROSITE: PS00165; DEHYDRATASE_SRP_THR; 1.
KW Isoleucine biosynthesis; Lyase; Chloroplast; Pyridoxal phosphate;
KW Allosteric enzyme; Transit peptide.
FT TRANSIT 1 51 CHLOROPLAST.
FT CHAIN 52 595 THREONINE DEHYDRATASE BIOSYNTHETIC.
FT BINDING 143 143 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 595 AA; 64937 MW; AC430B5DD9F0348 CRC64;

Query Match 20.9%; Score 363; DB 1; Length 595;
Best Local Similarity 32.1%; Pred. No. 2.3e-20;
Matches 96; Conservative 62; Mismatches 129; Indels 12; Gaps 8;

QY 26 TPVLTSSILNLTGRNLFKCELFQKSGFKIRGALNAVRLPYDALKERPKAVVTHSSG 85
Db 113 SPLELAEKLSDRIGVNFYIKREDKQKRVFKLGRYNNMNSNLSRELD---KGVITASAG 169

QY 86 NHGQALTYAAKLEGIPAYIVPOTAPDCKKLAIQAYCASIV-YCEPSDESRENVAKRUTE 144
Db 170 NHAQGVALAGQRLNCVAKIVMPTTTPQIKIDAVRALGGDVLYGKTFDEAQTHALE--LS 227

QY 145 ETEGI-MVHPNQBPVAVIAGQGTALVNLQVPLVDALVVPVGGGMLAGIAITVTKPKPS 203
Db 228 EKDLGYIIPFPDGPVIGKGGTIGTEINRQLKDIHAVFIPVGGGLIAGVATFFKQIAPN 287

QY 204 VKVYAAEPSNADCYOSKLGKLMPLNLYPETIADGVK-SSICLNTWPIIRDLVDDIFTV 262
Db 288 TKIIGVEPYGAASMTLSLHEGHRV-KLSNVDTFADGVAVALVGEYTFKQCQLIDGMVLY 346

QY 263 TEDEIKCATQLVWERMKLLIETAGVGAVALSQ-HFQVSPVSKNICIVLSGNVDLT 320
Db 347 ANDGISAAIKDYDEGRNILETSQAVAIAGAAAYCEFYKIKNE--NIVAIASGANMDFS 403
```

Search completed: June 24, 2003, 06:19:31

**THIS PAGE BLANK (USPTO)**







Db 400 AACTTTGATGAAGCGAAACGCAAGCGATGCGAAGTCTACAGCAGCAGGCGTTTCACCTCG 459  
QY 451 GTCCATCCCAACAGCAGGCTCGAGTGATAGTGGACAGGAACAATTGCCCTGGGAAGTG 510  
Db 460 GTGCCCGCTTGCACCATCCGATGCGATTGCGCGCAAGCGACGCTGGCGTGGAACTG 519  
QY 511 CTGAACGAGTTCCTTGTGTAGATGCACTGGTGTACAGTACAGGAGGAGGAGGATGGTT 570  
Db 520 CTCCACAGAGCCCATCTGACCCCGTATTGTGCCAGTGGCGCGCGGTCTGGCT 579  
QY 571 GCTGGAATAGCCATTACAATTAAAGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGA 629  
Db 580 GCTTGGTGGCGTGTGATCAACAACACTGATGCCCAACTCAAAAGTATGCGCGGTAGA 638

## RESULT 5

US-08-628-039-5  
; Sequence 5, Application US/08628039  
; Patent No. 5942660  
; GENERAL INFORMATION:  
; APPLICANT: Gruys, Kenneth J.  
; APPLICANT: Mitsky, Timothy A.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stark, David M.  
; APPLICANT: Hinchee, Maud A. W.  
; APPLICANT: Clemente, Thomas E.  
; APPLICANT: Connor-Ward, Dannelle V.  
; APPLICANT: Fedele, Mary J.  
; APPLICANT: Fry, Joyce E.  
; APPLICANT: Howe, Arlene R.  
; APPLICANT: Rozman, Renee J.  
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and  
; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalera  
; TITLE OF INVENTION: in Bacteria and Plants  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gary M. Bond, Monsanto Company, A3SB  
; STREET: 800 No. 5942660th Lindbergh Boulevard  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63167  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 13-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bond, Gary  
; REGISTRATION NUMBER: 29,283  
; REFERENCE/DOCKET NUMBER: 38-21(13585)A  
; TELEPHONE: (314)694-3412  
; TELEFAX: (314)695-5435  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1545 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)

US-08-628-039-5

Query Match 6.0%; Score 60.6; DB 2; Length 1545;  
Best Local Similarity 46.5%; Pred. No. 2.5e-10;  
Matches 195; Conservative 0; Mismatches 224; Indels 0; Gaps 0;  
QY 211 ACGCCAGAGAAGCCCAAGCCGTAGTACTACAGCAGCGAAGCAATTCCTGCTTACATGTGTTCCCAACAGCT 330  
Db 220 ACGGAAGAACAGAAAGCGCACGCGGTGATCACTGCTTCTGCGGGTAACACGCGCAGGGC 279  
QY 271 CTCACCTATGCTGCTAACTGGAAGAAATTCCTGCTTACATGTGTTCCCAACAGCT 330  
Db 280 GTGCGCTTTTCTTCGCGCGGTAGCGCTGAAGCGCTGATCGTTATGCCAACCGCCACC 339  
QY 331 CCCAACTGCAAGAAACTGGCAATCCAAAGCCTATGAGAGCATCGATATATATCTGTGACCCA 390  
Db 340 GCGGACATCAAGTCGACCGGTGCGCGGTTCGCGCGGGAAGTGTCTCCACGCGCG 399  
QY 391 AGTGACGAGTCCAGAGAAAGGTCACTCAAAAGAAATTTATGCAAGAAACAGAAAGCATCTTTG 450  
Db 400 AACTTTGATGAAGCGAAACGCAAGCGATGCGAAGTCTACAGCAGCAGGCGTTTCACCTCG 459  
QY 451 GTCCATCCCAACAGCAGGCTCGAGTGATAGTGGACAGGAACAATTCGCCCTGGGAAGTG 510  
Db 460 GTGCCCGCTTGCACCATCCGATGCGATTGCGCGCAAGCGACGCTGGCGTGGAACTG 519  
QY 511 CTGAACGAGTTCCTTGTGTAGATGCACTGGTGTACAGTACAGGAGGAGGAGGATGGTT 570  
Db 520 CTCCACAGAGCGCCCATCTCGACCCGTTATTGTGCCAGTGGCGCGCGGTCTGGCT 579  
QY 571 GCTGGAATAGCCATTACAATTAAAGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGA 629  
Db 580 GCTTGGTGGCGTGTGATCAACAACACTGATGCCCAACTCAAAAGTATGCGCGGTAGA 638

## RESULT 6

US-08-628-039-7  
; Sequence 7, Application US/08628039  
; Patent No. 5942660  
; GENERAL INFORMATION:  
; APPLICANT: Gruys, Kenneth J.  
; APPLICANT: Mitsky, Timothy A.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stark, David M.  
; APPLICANT: Hinchee, Maud A. W.  
; APPLICANT: Clemente, Thomas E.  
; APPLICANT: Connor-Ward, Dannelle V.  
; APPLICANT: Fedele, Mary J.  
; APPLICANT: Fry, Joyce E.  
; APPLICANT: Howe, Arlene R.  
; APPLICANT: Rozman, Renee J.  
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and  
; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyval  
; TITLE OF INVENTION: in Bacteria and Plants  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gary M. Bond, Monsanto Company, A3SB  
; STREET: 800 No. 5942660th Lindbergh Boulevard  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63167  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 13-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bond, Gary  
; REGISTRATION NUMBER: 29,283  
; REFERENCE/DOCKET NUMBER: 38-21(13585)A  
; TELEPHONE: (314)694-3412  
; TELEFAX: (314)695-5435  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1545 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)

US-08-628-039-5

```

; ADDRESSEE: Gary M. Bond, Monsanto Company, A3SB
; STREET: 800 No. 5942660th Lindbergh Boulevard
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63167
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,039
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE: 13-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bond, Gary
; REGISTRATION NUMBER: 29,283
; REFERENCE/DOCKET NUMBER: 38-21(13585)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)694-3412
; TELEFAX: (314)695-5435
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1545 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-628-039-8
;
; Query Match 6.0%; Score 60.6; DB 2; Length 1545;
; Best Local Similarity 46.5%; Pred. No.2.5e-10;
; Matches 195; Conservative 0; Mismatches 224; Indels 0; Gaps 0
;
; QY 211 ACGCCAGAGAGAGAGCCCAAGCCGTAGTTACTCACAGCAGCGAAACCATGGCCAAAGCT 270
; DB 220 ACGGAGACAGAGAAAGCGCAGCGGTGATCACTCTTCTGCGGTAAACCACGCGCAGGCG 279
; QY 271 CTCACCTTATGCTCTAAACTGGAAGAAATTCCTGCTTACATTTGGTTTCCCAACACAGCT 330
; DB 280 GTCGCGGTTTCTTCTGCGGTTAGGCGTGAAGCCCTGATCGTTATGCCAACGCCACC 339
; QY 331 CCCAACTGCAAGAACTGGCAATCCAGCCTATGGAGCATCGATGATATCTGTGACCCA 390
; DB 340 GCGGACATCAAGTCGACCGGCTTCGCGGCTTCGCGGCGGAAGTGTCTCCACGCGCG 399
; QY 391 AGTGACGAGTCCAGAGAAAAGGTCACTCAAGAATTTATGCAAGAAACAGAAAGCATCTTG 450
; DB 400 AACTTTGATGAAGGAAACCGCAAGCGATCGAATGTTCACAGCAGCGGGTTACCTTGG 459
; QY 451 GTCATATCCCAACAGGAGCGCTGCACTGATAGCTGGACAAGGAACAATTCGCTTGGAAAGT 510
; DB 460 GTCCGCGCGTTCGACCATCCGATGGTATTCGCGGCAAGGCAAGCTGCGCTGGAAGT 519
; QY 511 CTGAACCAAGTTTCCCTTGGTAGATGCACTGGTGTACAGTAGGAGGAGGAGGAATGTT 570
; DB 520 CTCACGAGGACGCCCATCTCGACCGGCTATTTGTGCGAGTCGCGGCGGCGGTCTGGCT 579
; QY 571 GCTGGAATACCCATTACAATTAAGGCCCTGAAACCTAGTGTGAAGGTATACGCTCTCA 629
; DB 580 GCTTGGCTGGCGGTGCTGATCAACAACAACTGATGCCGCAAAATCAAGTGATCGCGTGA 638
;
; RESULT 8
; US-08-673-388-1
; Sequence 1, Application US/08673388
; Patent No. 5958745
; GENERAL INFORMATION:
; APPLICANT: Gruy, Kenneth J.

```

APPLICANT: Mitsky, Timothy A.  
 APPLICANT: Kishore, Ganesh M.  
 APPLICANT: Slater, Steven C.  
 APPLICANT: Padgett, Stephen R.  
 APPLICANT: Stark, David M.  
 APPLICANT: Hinchee, Maud A. W.  
 APPLICANT: Clemente, Thomas E.  
 APPLICANT: Connor-Ward, Dannette V.  
 APPLICANT: Fedele, Mary J.  
 APPLICANT: Fry, Joyce E.  
 APPLICANT: Howe, Arlene R.  
 APPLICANT: Rozman, Renee J.  
 TITLE OF INVENTION: Methods of Optimizing Substrate Pools and  
 TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalerate  
 TITLE OF INVENTION: in Bacteria and Plants  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Charles E. Cohen, Monsanto Company, BB4F  
 STREET: 700 Chesterfield Parkway No. 5958745th  
 CITY: St. Louis  
 STATE: Missouri  
 COUNTRY: USA  
 ZIP: 63198  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/673,388  
 FILING DATE: 13-MAR-1996  
 CLASSIFICATION: 800  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Cohen, Charles E.  
 REGISTRATION NUMBER: 34,565  
 REFERENCE/DOCKET NUMBER: 38-21(13618)A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (314)537-6224  
 TELEFAX: (314)537-6047  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1545 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-673-388-1

```

Db      520 CTCAGCAGGAGCGCCATCTCGAGCGCGTATTGTGTCAGTGGCGGGCGGTCTGGCT 579
QY      571 GCTGGAATAGCCATTACAAATTAAGGCCCTGAAACCTAGTGTGAAGGTATACGCTGCTGA 629
Db      580 GCTTGGCGTGGCGTCTGATCAACAACACTGATGCCGCAATCAAAGTGTATCCCGTAGA 638

RESULT 9
US-08-673-388-5
; Sequence 5, Application US/08673388
; Patent No. 5958745
; GENERAL INFORMATION:
; APPLICANT: Gruys, Kenneth J.
; APPLICANT: Mitsky, Timothy A.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Padgett, Stephen R.
; APPLICANT: Stark, David M.
; APPLICANT: Hinchee, Maud A. W.
; APPLICANT: Clemente, Thomas E.
; APPLICANT: Connor-Ward, Dannette V.
; APPLICANT: Fedele, Mary J.
; APPLICANT: Fry, Joyce E.
; APPLICANT: Howe, Arlene R.
; APPLICANT: Rozman, Renee J.
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and
; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hy
; TITLE OF INVENTION: in Bacteria and Plants
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Charles E. Cohen, Monsanto Company, BB4F
; STREET: 700 Chesterfield Parkway No. 5958745th
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/673,388  
FILING DATE: 13-MAR-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Cohen, Charles E.  
REGISTRATION NUMBER: 34,565  
REFERENCE/DOCKET NUMBER: 38-21(13618)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314)537-6224  
TELEFAX: (314)537-6047  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1545 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-6733-388-5

	Query Match	5.0%;	Score 50.6;	DB 2:	Length 1545;
	Best Local Similarity	46.5%;	Pred. No. 2.5e-10;		
	Matches 195;	Conservative 0;	Mismatches 224;	Indels 0;	Gaps 0;
QY	211	ACGCCAAGAGAACCCAAAGCCGTAGTTACTCACAGCAGCGAAACCATGGCCAAAGCT	270		
Db	220	ACGSAAGACAGAAAGCGCACGGCGTGATCACTCTTCTGCGGGTAAACACACGCGCAGGCG	279		
QY	271	CTCACCTATGCTGCTAAACTGGAAGGAATTCCTGCTTACATGTTGGTTCCCAACAACAGCT	330		
Db	280	GTGCGGTTTCTCTGCGCGGTTAGGCGTGAAGCCCTGATCGTTATGCCAACGCGCAC	339		



```

: ADDRESSEE: Charles E. Cohen, Monsanto Company, BB4F
: STREET: 700 Chesterfield Parkway No. 5959179th
: CITY: St. Louis
: STATE: Missouri
: COUNTRY: USA
: ZIP: 63198
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA: US/08/614,877
: APPLICATION NUMBER: US/08/614,877
: FILING DATE: 13-MAR-1996
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Cohen, Charles E.
: REGISTRATION NUMBER: 34,565
: REFERENCE/DOCKET NUMBER: 38-21(10695)A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (314)537-6224
: TELEFAX: (314)537-6047
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1545 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
:
: US-08-614-877-1
:
: Query Match 6.0%; Score 60.6; DB 2: Length 1545;
: Best Local Similarity 46.5%; Pred. No.2.5e-10;
: Matches 195; Conservative 0; Mismatches 224; Indels 0; Gaps 0;
:
: QY 211 ACGCCAAGACAGAGCCCAAGCCGTAGTTACTCACAGACGCGAAACCATGGCCAAAGCT 270
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: DB 220 ACGAAGACAGAAAGCGCAGCGGCGTGATCACTGCTTCTCGCGGTATACCACGCCAGGGC 279
:
: QY 271 CTCACCTATGCTGCTAAACTGGAGGAATTCCTGCTTACATTTGGTTGCCCAACAGCT 330
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: DB 280 GTCCGCGTTTCTTCTGCGCGTTTAGGCGTGAAGGCCCTGATCGTTATGCCAACCGCCACC 339
:
: QY 331 CCCAACTGCAAGAACTGGCAATCCAAGCCCTATGGAGCATCGATAGTATACTGTGACCCA 390
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: DB 340 GCGGACATCAAGTCGACCGGCTGCGCGGCTTCGGCGGCGAAGTGTGCTCCACGGCGCG 399
:
: QY 391 AGTGACGAGTCCAGAGAAAAGGTGCATCAAGAAATATTGCAAGAAACAGAAAGCATCTTG 450
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: DB 400 AACTTTGATCAAGCGAAACGCAAAAGCGATCGAACTGTACAGCAGCAGGCGGTTTCACTGG 459
:
: QY 451 GTCCATCCCAACAGGAGCCCTGCAGTGTAGCTGGACAAGGAACAATTGCCCTGGAAGTG 510
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: DB 460 GTCCGCGCGTTGACCACTCCGATGGTGATTTGCCGGGAAGCAGCGCTGGCGCTGGAACGT 519
:
: QY 511 CTGAACCAAGGTTCCCTTTGGTAGATGCACCTGCTGTGTACCGTAGCAGGAGGAGGAATGTT 570
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: DB 520 CTCACGACGACGCCCATCTCGACCGCGTATTTGTGCCAGTCGCGCGGCGGCTCTGGCT 579
:
: QY 571 GCTGGAATAGCCATTACAATTAAGGCCCTGAAACCTAGTGTGAAGGTATACGCTGCTGA 629
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: DB 580 GCTTTGGTGGCGGTGCTGATCAACAACAACTGATGCCGAAATCAAAAGTGATCGCCGTAGA 638
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 400 AACCTTTGATGAAGCGAAGCAACCAAGCGATCGAACTGTCCACAGCAGCAGGGGTTCCACCTGG 459  
 Qy 451 GTCCATCCCAACAGGAGCGCTGCAGTGATAGCTGCAGCAAGGAAACAATTTGCCCTCGAAGTG 510  
 Db 460 GTGCGCGCTTGCACCATCCGATGGTGATTTGCCGGCAAGGCACGCTGGCGCTGGAACTG 519  
 Qy 511 CTGAACCACGAGTTCCTTTGGTAGATGCACCTGGTGTCACAGTAGGAGGAGGAGGAATGTT 570  
 Db 520 CTCACAGCAGCGCCCATCTCGACCGCGTATTTGTGCCAGCTCGCGCGCGCGTCTGGCT 579  
 Qy 571 GCTGGGAATAGCCATTACAATTAAGGCCCTGAAACCTAGTGTGAAGGTATACGCTGCCTGA 629  
 Db 580 GCTTGGCTGGCGGTGCTGATCAAAACAACCTGATGCCGCAAAATCAAAGTAGTCGCCGTAGA 638  
 RESULT 13  
 US-08-614-877-5  
 ; Sequence 5, Application US/08614877  
 ; Patent No. 5959179  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gruys, Kenneth J.  
 ; APPLICANT: Mitsky, Timothy A.  
 ; APPLICANT: Kishore, Ganesh M.  
 ; APPLICANT: Slater, Steven C.





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 24, 2003, 04:18:31 ; Search time 21.4683 Seconds  
(without alignments)  
464.608 Million cell updates/sec

Title: US-09-889-609B-8  
Perfect score: 1740  
Sequence: 1 MCAQYCISFADVEKAHINQ.....TSLNWVGQARPAPIQTQVSV 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/1aa/5A-COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/5B-COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/6A-COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/6B-COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS-COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1582.5	90.9	340	4	US-09-789-300A-2
2	365	21.0	424	4	US-09-134-001C-3876
3	359.5	20.7	441	1	US-08-403-866-10
4	303	17.4	436	3	US-08-669-378-2
5	303	17.4	436	3	US-08-669-378-12
6	302	17.4	436	3	US-08-669-378-4
7	302	17.4	436	3	US-08-669-378-6
8	302	17.4	436	3	US-08-669-378-10
9	300	17.2	436	3	US-08-669-378-8
10	296.5	17.0	378	4	US-09-789-300A-4
11	254.5	14.6	325	4	US-09-088-435-1
12	229	13.2	367	4	US-09-134-001C-4168
13	185.5	10.7	551	1	US-08-120-960-2
14	185.5	10.7	551	4	US-09-347-878-9
15	180	10.3	312	4	US-09-134-001C-3920
16	147.5	8.5	319	4	US-09-134-001C-3330
17	128.5	7.4	497	4	US-09-594-193-2
18	125	7.2	487	4	US-09-594-193-13
19	112	6.4	499	4	US-09-594-193-4
20	111.5	6.4	404	4	US-08-887-534A-34
21	104	6.0	434	4	US-09-594-193-7
22	101.5	5.8	484	4	US-09-066-046-8
23	93.5	5.4	403	4	US-09-134-001C-5236
24	93.5	5.4	540	4	US-08-973-462-22
25	93.5	5.4	1786	4	US-08-973-462-8
26	93	5.3	1114	2	US-08-576-626A-31
27	91.5	5.3	619	4	US-09-066-046-2

28	90.5	5.2	212	4	US-08-973-462-25	Sequence 25, Appl
29	90.5	5.2	609	4	US-08-980-115-11	Sequence 11, Appl
30	90.5	5.2	630	4	US-08-973-462-9	Sequence 9, Appl
31	88	5.1	947	4	US-09-228-986-73	Sequence 73, Appl
32	88	5.1	1416	1	US-08-061-465-4	Sequence 4, Appl
33	88	5.1	1642	2	US-08-662-227-2	Sequence 2, Appl
34	88	5.1	1642	4	US-09-017-947-2	Sequence 2, Appl
35	87	5.0	421	2	US-08-576-626A-53	Sequence 53, Appl
36	86	4.9	553	2	US-08-663-566A-13	Sequence 13, Appl
37	86	4.9	553	2	US-08-484-575A-14	Sequence 14, Appl
38	86	4.9	553	2	US-08-023-610-13	Sequence 13, Appl
39	86	4.9	553	2	US-08-288-065A-13	Sequence 13, Appl
40	86	4.9	553	2	US-08-362-240A-13	Sequence 13, Appl
41	86	4.9	553	3	US-08-477-459-14	Sequence 14, Appl
42	86	4.9	553	3	US-08-479-869-14	Sequence 14, Appl
43	86	4.9	553	4	US-08-486-414-14	Sequence 14, Appl
44	86	4.9	553	4	US-08-804-372A-11	Sequence 11, Appl
45	86	4.9	553	5	PCT-US94-01826A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1

US-09-789-300A-2

; Sequence 2, Application US/09789300A

; Patent No. 6458576

; GENERAL INFORMATION:

; APPLICANT: Meyers, Rachel

; APPLICANT: Rudolph-Owen, Laura A.

; TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate

; FILE OF INVENTION: Dependent-Enzyme Family Member and Uses Therefore

; FILE REFERENCE: 35800/208926

; CURRENT APPLICATION NUMBER: US/09/789,300A

; CURRENT FILING DATE: 2001-02-20

; PRIOR APPLICATION NUMBER: US 60/183,208

; PRIOR FILING DATE: 2000-02-17

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 340

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-09-789-300A-2

Query Match 90.9%; Score 1582.5; DB 4; Length 340;

Best Local Similarity 89.7%; Pred. No. 9.7e-168;

Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;

Qy 1 MCAQYCISFADVEKAHINQDSIHLPVLTSSILNQIAGRNLFFKCELFQKTSFKIRGA 60

Db 1 MCAQYCISFADVEKAHINRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60

Qy 61 LNAIRGLIPDPEKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPTAPNCKKLAIOA 120

Db 61 LNAVRSVLPDALERKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPTAPDCKKLAIOA 120

Qy 121 YGASIVYCDPDSREKVTQRIQIMQTEGILVHPNQEPAVIAGOGTIALEVLNQPLVDAL 180

Db 121 YGASIVYCDPDSREKVTQRIQIMQTEGILVHPNQEPAVIAGOGTIALEVLNQPLVDAL 180

Qy 181 VVPVGGGGMVAGIATITKALKPSVKVYAAEPNADDCYQSKLGBELTNLHPPEIADGV 240

Db 181 VVPVGGGGMVAGIATITKALKPSVKVYAAEPNADDCYQSKLGBELTNLHPPEIADGV 240

Qy 241 KSSICLNTWPIIRLDVDDFTVTEDEIKATQLVWGRMKLLIEPTAGVALAVALSHQFQT 300

Db 241 KSSICLNTWPIIRLDVDDFTVTEDEIKATQLVWGRMKLLIEPTAGVALAVALSHQFQT 300

Qy 301 VSPEVKNCIVLSGNGVDLT-SLNVGQARPAPIQTQVSV 339

Db 301 VSPEVKNCIVLSGNGVDLTSSITWVQKQARPAPIQTQVSV 340

NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30, 727  
REFERENCE/DOCKET NUMBER: 20747/30  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1600  
TELEFAX: (716) 263-1487  
TELEX: 978450 (WUT)  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 441 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Lactococcus lactis subsp. lactis  
INDIVIDUAL ISOLATE: ILVA  
US-08-403-866-10

Query Match 20.7%; Score 359.5; DB 1;  
Best Local Similarity 29.4%; Pred. No. 2,9e-31;  
Matches 98; Conservative 67; Mismatches 145; Indels 23; Gaps 3

QY	14	KAHINIQDSIHLPVUTSSILNQIAGRNLIFFKCELFQKTSFKIRGALNAIRGLIPDTP	73
Dd	34	QANIYLKEVTPTQLQDPYLSKNKYQANTYLYKEENLQKVRSPFLRGAYYSISKL---	90
QY	74	EXPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPNCKKLIAOYGASIVVC---	129
Dd	91	ORSKGWCVCSAGNHAGVAFANQLNISATIFMPVTPPNOKISOVKRPFEGESHVTIRLIGD	150
QY	130	PDESREKVTRIMQETEGILVHPNOEPAVIAGOGTIALEVLNQP-----LVDAVVPPVG	185
Dd	151	TDES-ARAKAFSQNDXFPIDPDDENVIAGOGTVALEIFAQAOKKGISLDKIFVOIC	209
QY	186	GGGWAGIAITIKALKPSVKVYAAEPSNADCCVOSKLGELTPLNLHPPETIADGVK--SSI	244
Dd	210	GGGLIAGITAYSERYPQTEIIIGVEAKGATSMKAAYSAGO-PVTFLEHDKFADGI VATV	268
QY	245	GLNTWPIIRDVDVFVTEDETKYATOLVWGMRMKLLIEPTAGVALAAYLSQHFOIVSPE	304
Dd	269	GOKTYOLIINDKKVQLLAVIDEGLISOTILEYSLKGI VAEPAGATSVAAL-----ELIKDE	323
QY	305	V--KNVICVLSGCNVDLTSLNWVGQAEPAPYQ	335
Dd	324	IKGKNIVCIISGCNDISRMQEII--EERALVYE	354

RESULT 4  
US-08-669-378-2

7 PAGES: NO. 6107/003  
 7  
 7 GENERAL INFORMATION:  
 7  
 7 APPLICANT: Moeckel, Bettina  
 7 APPLICANT: Eggeling, Lothar  
 7 APPLICANT: Sahm, Hermann  
 7  
 7 TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF  
 7 TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE  
 7 TITLE OF INVENTION: DEHYDRATASE  
 7  
 7 NUMBER OF SEQUENCES: 18  
 7  
 7 CORRESPONDENCE ADDRESS:  
 7  
 7 ADDRESSEE: Foley & Lardner  
 7 STREET: 3000 K Street, N.W., Suite 500  
 7 CITY: Washington  
 7 STATE: D. C.  
 7 COUNTRY: USA  
 7  
 7 ZIP: 20007-5109  
 7  
 7 COMPUTER READABLE FORM:  
 7  
 7 MEDIUM TYPE: Floppy disk  
 7 COMPUTER: IBM PC compatible  
 7 OPERATING SYSTEM: PC-DOS/MS-DOS  
 7 SOFTWARE: Patent In Release #1.0, Version #1.30  
 7  
 7 CURRENT APPLICATION DATA:  
 7

APPLICATION NUMBER: US/08/669,378  
FILING DATE: 20-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DE95/00017  
FILING DATE: 09-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 00 926.7  
FILING DATE: 14-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 016881/0142  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-669-378-2

Query Match 17.4%; Score 303; DB 3; Length 436;  
Best Local Similarity 28.1%; Pred. No. 5.7e-25;  
Matches 94; Conservative 60; Mismatches 159; Indels 22; Gaps 10;  
QY 7 ISFADVEKAHINQDSIHHTPVLTSSILNQIAGRNLFKCELFQKTSKIRGALNAIRG 66  
DB 21 IRAADQTAQARISSVIAPTPLQYCPRLSEETGAETIYLRKREDLDQVRSYKIRGALNS --- 77  
QY 67 LIPDTPPEPKPA-VVTHSSGNHQALTYAAKLEGIPAYIVWPQTAPNCKKLAIOAYG --- 122  
DB 78 -GAQSPQEQORDAGIVAASAGNHAQVAYVCKSLGVQGRIVVPVQPKQRDRIMVHGGEF 136  
QY 123 ASIVCYDPSDESEKREKVTQRMQETEGILVHPNOEPVAVIAGQGTIALEVLNQVPLV ---D 178  
DB 137 VSLVVTGNFNDEASAAAHEDAERTGATLTIEPEDARNTVIGQGTVAEAILSQLTSMGKSAD 196  
QY 179 ALVVPVGGGGMVAGIATITKALKPSVKVYAAPSNAADDQYQSKLKGELTPNLHPETIAD 238  
DB 197 HWVPVGGGLLAGVVSVMADMAPRTAIVGIEPAGAAS-MQALHNGGPITLTETVDPFVD 255  
QY 239 GVK-SSIGLNTWPIIRDLDVDDVFTVTEDEIKYATOL--VWGRMKLLIEPTAGVALAAVLS 295  
DB 256 GAEVKRVGDLNVTIVEKNOGRVHMMSATGAVCTEMLDLYQNEGIIAEPAAGALSIALKE 315  
QY 296 QHFQTVSPVKVNCIVLSGNGVDLTSLNHWGOAER 330  
DB 316 MSF---APGSVVVCII-SGGNNDV--LRYAEIAER 344

RESULT 5  
US-08-669-378-12  
Sequence 12, Application US/08669378  
Patent No. 6107063  
GENERAL INFORMATION:  
APPLICANT: Moekkel, Bettina  
APPLICANT: Eggeling, Lothar  
APPLICANT: Sahm, Hermann  
TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF  
TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE  
TITLE OF INVENTION: DEHYDRATASE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/669,378  
FILING DATE: 20-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DE95/00017  
FILING DATE: 09-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 00 926.7  
FILING DATE: 14-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 016881/0142  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-669-378-12

Query Match 17.4%; Score 303; DB 3; Length 436;  
Best Local Similarity 28.4%; Pred. No. 5.7e-25;  
Matches 95; Conservative 59; Mismatches 159; Indels 22; Gaps 10;  
QY 7 ISFADVEKAHINQDSIHHTPVLTSSILNQIAGRNLFKCELFQKTSKIRGALNAIRG 66  
DB 21 IRAADQTAQARISSVIAPTPLQYCPRLSEETGAETIYLRKREDLDQVRSYKIRGALNS --- 77  
QY 67 LIPDTPPEPKPA-VVTHSSGNHQALTYAAKLEGIPAYIVWPQTAPNCKKLAIOAYG --- 122  
DB 78 -GAQSPQEQORDAGIVAASAGNHAQVAYVCKSLGVQGRIVVPVQPKQRDRIMVHGGEF 136  
QY 123 ASIVCYDPSDESEKREKVTQRMQETEGILVHPNOEPVAVIAGQGTIALEVLNQVPLV ---D 178  
DB 137 VSLVVTGNFNDEASAAAHEDAERTGATLTIEPEDARNTVIGQGTVAEAILSQLTSMGKSAD 196  
QY 179 ALVVPVGGGGMVAGIATITKALKPSVKVYAAPSNAADDQYQSKLKGELTPNLHPETIAD 238  
DB 197 HWVPVGGGLLAGVVSVMADMAPRTAIVGIEPAGAAS-MQALHNGGPITLTETVDPFVD 255  
QY 239 GVK-SSIGLNTWPIIRDLDVDDVFTVTEDEIKYATOL--VWGRMKLLIEPTAGVALAAVLS 295  
DB 256 GAEVKRVGDLNVTIVEKNOGRVHMMSATGAVCTEMLDLYQNEGIIAEPAAGALSIALKE 315  
QY 296 QHFQTVSPVKVNCIVLSGNGVDLTSLNHWGOAER 330  
DB 316 MSF---APGSVVVCII-SGGNNDV--LRYAEIAER 344

RESULT 6  
US-08-669-378-4  
Sequence 4, Application US/08669378  
Patent No. 6107063  
GENERAL INFORMATION:  
APPLICANT: Moekkel, Bettina  
APPLICANT: Eggeling, Lothar  
APPLICANT: Sahm, Hermann  
TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF  
TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE  
TITLE OF INVENTION: DEHYDRATASE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,378
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DE95/00017
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 00 926.7
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016881/0142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-669-378-4

Query Match 17.4%; Score 302; DB 3; Length 436;
Best Local Similarity 28.1%; Pred. No. 7.3e-25;
Matches 94; Conservative 60; Mismatches 159; Indels 22; Gaps 10;

QY 7 ISFADVEKAHINIQDSIHLPVLTSSILNQIAGRNLFFKCELFQKTSFKIRGALNAIRG 66
Db 21 IRAADIQTAAQARISSVIAPTPLQYCPRLSEETGAETYLKREDLQDVRSYKIRGALNS--- 77
QY 67 LIPDTPPEPKPA-VVTHSSGNHGQALTYAAKLEGIPAYIVVPTQAPNCKKLAIQAYG--- 122
Db 78 -GAQSPQEQORDAGIVAASAGNHAQGVAYVCKSLGVQGRIVVPVQTPKQKRDRI MVHGGEF 136
QY 123 ASIVYCDPSDESREKVTQRMQETEGILVHPNQEPAVIAGOGTIALEVLNQVPLV---D 178
Db 137 VSLVVTGNFDEASAAAHEDAERTGATLIEPFDARTNTVIGQGTVAEILSQTSMKSD 196
QY 179 ALVVPVGGGMMVAGIATIKALKPSVKVYAAPSNADDCYQSKLKGELTPNLPHPPTIAD 238
Db 197 HVMVPVGGGGLLAGVSVYVYADNAPRTAIVGIEPAGAAS-MQAALHNGGPTITLETVDPPVD 255
QY 239 GVK-SSIGLNTWPIIRDLVDVFTTEDEIKYATOL--VWGRMKLLIEPTAGVALAULS 295
Db 256 GAEVRRVGLNNTIVTEKNGRVHMSATEGAVCTEMLDLYQNEGIATPAGALSAGLKE 315
QY 296 QHFQTVSPVKVNCVILSGNVDLTSLNWVGQAE 330
Db 316 MSF---APGSVVVCII-SGGNNDV--LRYAETAER 344

RESULT 7
US-08-669-378-6
; Sequence 6, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:
; APPLICANT: Moeckel, Bettina
; APPLICANT: Eggeling, Lothar
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,378
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DE95/00017
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 00 926.7
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016881/0142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-669-378-4

Query Match 17.4%; Score 302; DB 3; Length 436;
Best Local Similarity 28.1%; Pred. No. 7.3e-25;
Matches 94; Conservative 60; Mismatches 159; Indels 22; Gaps 10;

QY 7 ISFADVEKAHINIQDSIHLPVLTSSILNQIAGRNLFFKCELFQKTSFKIRGALNAIRG 66
Db 21 IRAADIQTAAQARISSVIAPTPLQYCPRLSEETGAETYLKREDLQDVRSYKIRGALNS--- 77
QY 67 LIPDTPPEPKPA-VVTHSSGNHGQALTYAAKLEGIPAYIVVPTQAPNCKKLAIQAYG--- 122
Db 78 -GAQSPQEQORDAGIVAASAGNHAQGVAYVCKSLGVQGRIVVPVQTPKQKRDRI MVHGGEF 136
QY 123 ASIVYCDPSDESREKVTQRMQETEGILVHPNQEPAVIAGOGTIALEVLNQVPLV---D 178
Db 137 VSLVVTGNFDEASAAAHEDAERTGATLIEPFDARTNTVIGQGTVAEILSQTSMKSD 196
QY 179 ALVVPVGGGMMVAGIATIKALKPSVKVYAAPSNADDCYQSKLKGELTPNLPHPPTIAD 238
Db 197 HVMVPVGGGGLLAGVSVYVYADNAPRTAIVGIEPAGAAS-MQAALHNGGPTITLETVDPPVD 255
QY 239 GVK-SSIGLNTWPIIRDLVDVFTTEDEIKYATOL--VWGRMKLLIEPTAGVALAULS 295
Db 256 GAEVRRVGLNNTIVTEKNGRVHMSATEGAVCTEMLDLYQNEGIATPAGALSAGLKE 315
QY 296 QHFQTVSPVKVNCVILSGNVDLTSLNWVGQAE 330
Db 316 MSF---APGSVVVCII-SGGNNDV--LRYAETAER 344

RESULT 8
```

```
US-08-669-378-10
; Sequence 10, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:
; APPLICANT: Moekkel, Bettina
; APPLICANT: Eggeling, Lothar
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
; TITLE OF INVENTION: DEHYDRATASE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,378
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DE95/00017
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 00 926.7
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016881/0142
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-669-378-10

Query Match 17.4%; Score 302; DB 3; Length 436;
Best Local Similarity 28.1%; Pred. No. 7.3e-25;
Matches 94; Conservative 60; Mismatches 159; Indels 22; Gaps 10;

QY 7 ISFADVEKAHINIQDSIHLPVTSSILNQIAGRNLFKCELFQKGTSGFKIRGALNAIRG 66
Db 21 IRAADIQTAAQARISSVIAPTPLQYCPRLSEETGAETIYLKREDLQDVRSYKIRGALNS--- 77

QY 67 LIPDTPPEPKA-VVTHSSGNHGOALTYAAKLEGIPAYIVVPTAPNCKKLAIQAYG--- 122
Db 78 -GAQSPQEQORDAGIVAASAGNAQGVAYVCKSLGVQGVRIYVVPQTPKQRDRIMVHGGEF 136

QY 123 ASIVYCDPDESREKVTQRIMQETEGILVHPNQEPAVIAGQGTIALEVLNQPLV----D 178
Db 137 VSLVVTGNFNDEASAAHAEDAERTGATLIEPFDARNTVIGQGTVAEIIQSLSMCKSAD 196

QY 179 ALVVPVGGGWMVAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIAD 238
Db 197 HVMPVPGGGLLAGVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 255

QY 239 GVK-SSIGLNTWPIIRLDVDDFTVTDEIKYATQL--VWGRMKLLIETAGVALAALVS 295
Db 256 GAEVKRVGDLNITVEKNOGRVHMSGATEGVCTEMLDLDLYONEGIIAEAPAGALSIALKE 315

US-08-669-378-8
; Sequence 8, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:
; APPLICANT: Moekkel, Bettina
; APPLICANT: Eggeling, Lothar
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
; TITLE OF INVENTION: DEHYDRATASE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,378
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DE95/00017
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 00 926.7
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016881/0142
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-669-378-8

Query Match 17.2%; Score 300; DB 3; Length 436;
Best Local Similarity 28.9%; Pred. No. 1.2e-24;
Matches 97; Conservative 58; Mismatches 157; Indels 24; Gaps 11;

QY 7 ISFADVEKAHINIQDSIHLPVTSSILNQIAGRNLFKCELFQKGTSGFKIRGALNAIRG 66
Db 21 IRAADIQTAAQARISSVIAPTPLQYCPRLSEETGAETIYLKREDLQDVRSYKIRGALNS--- 77

QY 67 LIPDTPPEPKA-VVTHSSGNHGOALTYAAKLEGIPAYIVVPTAPNCKKLAIQAYG--- 122
Db 78 -GAQSPQEQORDAGIVAASAGNAQGVAYVCKSLGVQGVRIYVVPQTPKQRDRIMVHGGEF 136

QY 123 ASIVYCDPDESREKVTQRIMQETEGILVHPNQEPAVIAGQGTIALEVLNQPLV----D 178
Db 137 VSLVVTGNFNDEASAAHAEDAERTGATLIEPFDARNTVIGQGTVAEIIQSLSMCKSAD 196

QY 179 ALVVPVGGGWMVAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIAD 238
Db 197 HVMPVPGGGLLAGVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 255

QY 239 GVK-SSIGLNTWPIIRLDVDDFTVTDEIKYATQL--VWGRMKLLIETAGVALAALVS 295
Db 256 GAEVKRVGDLNITVEKNOGRVHMSGATEGVCTEMLDLDLYONEGIIAEAPAGALSIALKE 315
```

Db 197 HVMVPVGGGLLAGVWSYMDAPRTAIVGIEPAGAS-MQAALHNGGPITLETVPDFVD 255  
QY 239 G--VKSSIGLNTWPIIRDLDVDTFTVEDEIKYATOL--VWGRMKLLIEPTAGVALAVAL 294  
Db 256 GGEVKRVGDJN--YTVIEKNGRVHMSATEGAVCTEMLDIYQNEGIIAEPAGALSTAGLK 314  
QY 295 SQHFTQVSPVKNCVIVLSGNNVDLTLNVLVWGQAE 330  
Db 315 EMSF---APGSVVVCI--SGGNDV--LRYAEIAER 344  
RESULT 10  
US-09-789-300A-4  
Sequence 4, Application US/09789300A  
Patent No. 6458576  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel  
APPLICANT: Rudolph-Owen, Laura A.  
TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate  
TITLE OF INVENTION: Dependent Enzyme Family Member and Uses Therefore  
FILE REFERENCE: 35800/208926  
CURRENT APPLICATION NUMBER: US/09/789,300A  
CURRENT FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: US 60/183,208  
PRIOR FILING DATE: 2000-02-17  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 378  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Pyridoxal-Phosphate Dependent Enzyme Family Domain  
US-09-789-300A-4

Query Match 17.0%; Score 296.5; DB 4; Length 378;  
Best Local Similarity 30.2%; Pred. No. 2.4e-24;  
Matches 115; Conservative 51; Mismatches 128; Indels 87; Gaps 17;  
QY 19 IQDSIHLPVLTSSILNQIAGRL-----FFKCE-LFQKGTGSPKIRG-ALNAI----- 64  
Db 1 VTELGNTPLVRLNLSKELGEGLGANAAVEIYKLEDLNGPTGSPKDRGLNLMILLAE 60  
QY 65 -----RGLIPDTP-EKPRAVVTHSSNGHQALTYAAKLEGIPAYIVVPTAPNCKKLA- 117  
Db 61 KLGKGGIVPTQVQVSKTTIIEPTSGNTGIALALAAALLGLKCTIVMPATDTSREKRAQ 120  
QY 118 IQAYGASIVCDPSDESREKVTQIRIMQETEGILVHPNQEPV-----IA 161  
Db 121 LRALGAELVVVPVAGGSDDLADAIAKAE--LAEENPENAYLLNQAAQGFDPNANPEIA 178  
QY 162 GQGTIALEVLNQV-----PLVDALVVPVGGGVMVAGIATITKALKPS-----V 204  
Db 179 GQRTIGPEIWEQGGKEISIGRLPDVAVPVGGGGTITGIARYLKLNPDKGIDVLELPV 238  
QY 205 KYAAEPSNADDCYQSKLKGELT---PN-LH-----PPEIADGVKSSIGLN-- 247  
Db 239 KVIQVPEGS-AVLSGSLKATLTLAGKPGPLHGRDSKYLLODEPVTLPETKSGIGLGV 297  
QY 248 -----TWPIIROL-----VDDVFTVTEDEIKYATQVLMGRMKLLIEPTAGVALAVALSQ 296  
Db 298 RVGEFVPPILDELLDRROGIDEVVTDEALEAARLLAREEGILVGPSSGAVAALKL 357  
QY 297 HFQTVSP--EVKNVCIVLSG 315  
Db 358 AKEGKKPLNKGKTIIVILSGG 378

RESULT 11  
US-09-088-435-1  
Sequence 1, Application US/09088435  
Patent No. 6277619

GENERAL INFORMATION:  
APPLICANT: LAL, PREETI  
APPLICANT: CORLEY, NEIL C.  
APPLICANT: GUEGLER, KARL J.  
APPLICANT: PATTERSON, CHANDRA  
TITLE OF INVENTION: SERINE DEHYDRATASE HOMOLOG  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/088,435  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Cerrone, Michael C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0512 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 325 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: TPL1A2S08  
CLONE: 2752518  
US-09-088-435-1  
Query Match 14.6%; Score 254.5; DB 4; Length 325;  
Best Local Similarity 28.1%; Pred. No. 8.8e-20;  
Matches 92; Conservative 59; Mismatches 117; Indels 59; Gaps 15;  
QY 20 QDSIH-LTPVLTSILNQIAGRLPFKCELFQKGTGSPKIRGALNAILGLIPDTPPEKPKA 78  
Db 11 QEPFHVVTPLLESWALSQVAGMPVFLKCNVQPSGSKIRG-----TGHFCQEMAKKGRH 66  
QY 79 VYTHSSNGHQALTYAAKLEGIPAYIVVPTAPNCKKLAIQAYGASIVYC-----DPSDES 134  
Db 67 LVCSSCGNAGIAAAYAAKLGIPATIVLPSTSLQVVQRLQSGAEVQLTGKVMDEANLR 126  
QY 135 REKVTQIRIMQETEGILVHPNQEPVAGGTGTALE---VLNQVPLVDALVVPVGGGMVA 191  
Db 127 AQELAKRQDGNW---VPPFDHPLIWKHSLVQELKAVLRTPP--GALVLAIVGGGGLLA 180  
QY 192 GTAITIKAL-----KPSKVYAAEPSNADDCYQSKL-KGELTPNLHPPETIAD--GVKSSI 244  
Db 181 GVVAGLLEVGWQHVPIAMHGAH-----CFNNAITAGKLV-----TLPDITSVAKSL 227  
QY 245 GLNT-----WPIIRDLDVDTFTVEDEIKYATQVLMGRMKLLIEPTAG-VALAA 292  
Db 228 GAKTVAARALECMQVCKIHSEVVEDTEAVS-----AVQOLLDDERMLVEPACGAAAIYS 281  
QY 293 VLSQHPQT---VSPEVKNCIVLSG 316  
Db 282 GLLRLQAEGLPPSLTSVVIVCGN 308

RESULT 12  
US-09-134-001C-4168.  
; Sequence 4168, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4168  
; LENGTH: 367  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4168

Query Match 13.2%; Score 229; DB 4; Length 367;  
Best Local Similarity 26.0%; Pred. No. 7.5e-17;  
Matches 86; Conservative 66; Mismatches 147; Indels 32; Gaps 12;  
QY 26 TPVLTSSILNQIAGRNLFFKCELFQKTSFKIRGALNAIRGLIPDTPPEKPKAVVTHSSG 85  
DB 44 TPLIYCNLSQKLDIELYVKEGANTGSEFKDRGMWA-----VTKAKEQKGVVCASTG 99  
QY 86 N-HQALTYAAKLEGIPAYIVVQTPAPNCKKLAIAQA--YGASTVYCDPSDESREKVTQRI 142  
DB 100 NTSASAAVAAR-AGLKAIVPEGKIALGKLS-QAVMYGAETVSTIEGNFDEALEIVKEI 157  
QY 143 MQETGILVHPNQPFAVIAQOGTIALEVLNQV--PLVDALVVPVGGGMVAG-----IAI 195  
DB 158 AEENDEIELVNSVNPRIEKGQKTAPEIVQLDQAPDILAIPVGNAGNITAYWQGFVEY 217  
QY 196 TIKALPSPVYAAEPSNADDCYQSKLKGELTNLHPPEIADGVKSSIG-LNTWPIIRD 254  
DB 218 HNKNTQLQPMQFGQFQAGASPIVQNKI-----IKNPETIATAIR--IGNPASWQKAVN 268  
QY 255 LVDD----VFTVTEDEIKYATQLVWGRMKLLIETAGVALAAVLSQHFQTVSPVKNVCI 310  
DB 269 ALDESNGLDSVIDEILEAYQLMTTNEGVSFEPASNASIAGLIKLRSGKLPKGGKIVA 328  
QY 311 VLSGGNV--DLTSLNWNQARPAFYQTVSV 339  
DB 329 ILTGNGLKDPDTAISLLDNPLOPLPNKESI 359

RESULT 13  
US-08-120-960-2  
; Sequence 2, Application US/08120960  
; Patent No. 5523225  
; GENERAL INFORMATION:  
; APPLICANT: KRAUS, JAN P  
; TITLE OF INVENTION: DNA SEQUENCE ENCODING HUMAN  
; TITLE OF INVENTION: CYSTATHIONINE B-SYNTHASE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DILWORTH & BARRESE  
; STREET: 4350 LA JOLLA VILLAGE DRIVE, SUITE 300  
; CITY: SAN DIEGO  
; STATE: CALIFORNIA  
; COUNTRY: U.S.A.  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/120,960  
; FILING DATE: 12-SEP-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PEPPER PH.D., FREDERICK W.  
; REGISTRATION NUMBER: 31,286  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-4410  
; TELEFAX: 619-453-2839  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 551 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-120-960-2

Query Match 10.7%; Score 185.5; DB 1; Length 551;  
Best Local Similarity 27.4%; Pred. No. 1e-11;  
Matches 81; Conservative 43; Mismatches 133; Indels 39; Gaps 10;  
QY 26 TPVLTSSILNQIAGR-----NLFFKCELFQKTSFKIRGALNAIRGLIPDTPPEKPKAVV 80  
DB 87 TPMVR---INKIGKFKLGKCELLAKCEFFNAGGSVKDRISLRMIEDAERDGTILKPGDTII 143  
QY 81 THSSGNHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIAQAYGASIVVCD-----PSDES 135  
DB 144 EPTSGNTGTLAALAAVAVRGYRCIIIVPEKMSSEKVDVLRALGAIEIVRPTNARFDPESH 203  
QY 136 EKVQIRIMQ--TEGILVHPNQPFAVIAQOGTIALEVLNQV--LVDAVVPVGGGMVAG 192  
DB 204 VGVAWRLKNEIPNSHILDOVRNASNPLAHYDTTADETLQCCDKGLMDLVAASVGTGGITG 263  
QY 193 IAITIKALPSPVYAAEPSNADDCYQSKLKGELTNLHPPEIADGVKSS-----IGLN 247  
DB 264 IARKLEKPCGRIIGVDP-----EGSI---LAEPPELNTEQTYTEVEGYID 309  
QY 248 TWPILI--RLVDVDFVTEDEIKYATQLVWGRMKLLIETAG--VALAAVLSQHFQ 299  
DB 310 FIPVLDRTVVDVWFKNSNDEEAFTRMIAQOGLCGGSAGSTVAVAVAAQELQ 365

RESULT 14  
US-09-347-878-9  
; Sequence 9, Application US/09347878C  
; Patent No. 6376210  
; GENERAL INFORMATION:  
; APPLICANT: Yuan, Chong  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES  
; FILE REFERENCE: 25885-1651  
; CURRENT APPLICATION NUMBER: US/09/347,878C  
; CURRENT FILING DATE: 1999-07-06  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 551  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-347-878-9

Query Match 10.7%; Score 185.5; DB 4; Length 551;  
Best Local Similarity 27.4%; Pred. No. 1e-11;  
Matches 81; Conservative 43; Mismatches 133; Indels 39; Gaps 10;  
QY 26 TPVLTSSILNQIAGR-----NLFFKCELFQKTSFKIRGALNAIRGLIPDTPPEKPKAVV 80  
DB 87 TPMVR---INKIGKFKLGKCELLAKCEFFNAGGSVKDRISLRMIEDAERDGTILKPGDTII 143  
QY 81 THSSGNHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIAQAYGASIVVCD-----PSDES 135  
DB 144 EPTSGNTGTLAALAAVAVRGYRCIIIVPEKMSSEKVDVLRALGAIEIVRPTNARFDPESH 203

```

136  QY  136  EKVQTRIMQE--TEGIIVHPNQEPAVIAGQGTIALEVLNQVP-LVDALVVPVGGGMVAG 192
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
204  Db  204  VGVAWRLKNEIPNSHILDQYRNASNPPLAHYDTTADETILQQCDGKLDMLVASVGTGTTG 263
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY    193  IAITIKALKPSVKVYAAEPSNADCYQSKLKGELTPNLHPPETIADGVKSS-----ICLN 247
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    264  IARKLKEKPCGCRRIIGVDP-----EGSI---LAEPEELNQETTTVEVEGIGYD 309
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY    248  TWPII---RDVLVDDVFTTVEDEIKYATOLVGMGRMKLLIETPTAG--VALAAVLSQHFQ 299
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    310  FIPTVLDRTVVDKWFKNDEEAFTFARMLIQAEGLLCGGSAGSTVAVAKAAQELQ 365
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 15
US-09-134-001C-3920
: Sequence 3920, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GPC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3920
: LENGTH: 312
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3920

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 07:02:30 ; Search time 1462 Seconds  
(without alignments)  
11332.410 Million cell updates/sec

Title: US-09-889-609B-9  
Perfect score: 1023  
Sequence: 1 atgtgtctcagttatgcattcat.....atcagtcgtttctgtttaa 1023

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	803	78.5	1398	11 AK017235	AK017235 Mus muscu
2	751	73.4	785	13 BI763753	BI763753 603047780
3	729.2	71.3	848	9 AU131397	AU131397 AU131397
4	602	58.8	634	14 BM719814	BM719814 UI-E-EJ0-
5	526	51.4	527	9 AA446793	AA446793 zw89f02.r
6	523.8	51.2	769	12 BF163794	BF163794 601769701

7	491.8	48.1	667	10	BB609829
8	487.6	47.7	652	10	AV709023
9	479.8	46.9	600	13	BI988879
10	479.2	46.8	697	10	AV649313
11	451	44.1	819	13	BI739030
12	450.8	44.1	565	10	BE655084
13	410.4	40.1	1159	14	BQ215405
14	402.4	39.3	918	14	BQ885621
15	391.8	38.3	823	13	BI738966
16	383	37.4	391	14	BQ303565
17	380.6	37.2	404	14	BQ303568
18	379.6	37.1	386	14	BQ303566
19	376.6	36.8	698	9	AV022510
20	365.4	35.7	455	9	AI322578
21	363.8	35.6	602	10	AV374491
22	352.8	34.5	471	14	H73097
23	348.6	34.1	729	13	BC965678
24	341	33.3	394	10	AV649514
25	337	32.9	793	12	BG404240
26	329	32.2	742	9	AI525507
27	328.2	32.1	827	13	BI332919
28	323.8	31.7	438	14	H86748
29	321.2	31.4	519	10	BE641631
30	320.8	31.4	830	12	BF135412
31	315.4	30.8	556	10	BB692486
32	300.2	29.3	380	9	AA034539
33	299	29.2	661	10	BB621410
34	295.4	28.9	688	10	BB625125
35	289.8	28.3	1087	12	BF532321
36	282	27.6	544	9	AA432108
37	275.2	26.9	645	10	AW211242
38	274.6	26.8	646	10	BB644584
39	274	26.8	566	14	BM766019
40	273	26.7	288	14	243188
41	266.4	26.0	632	10	BB641468
42	266	26.0	611	10	BB618702
43	265.4	25.9	657	10	AW319255
44	258.2	25.2	672	10	BE380996
45	255.6	25.0	426	14	D53895

## ALIGNMENTS

RESULT 1  
AK017235  
LOCUS AK017235 1398 bp mRNA linear HTC 19-JAN-2002  
DEFINITION Mus musculus adult male pituitary gland cDNA, RIKEN full-length enriched library, clone:5330405D10:serine racemase, full insert sequence.  
ACCESSION AK017235  
VERSION AK017235.1 GI:12856372  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (strain:C57BL/6J) adult male pituitary gland cDNA to mRNA, clone\_lib:RIKEN full-length enriched mouse cDNA library  
clone:5330405D10.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1  
AUTHORS Carninci,P. and Hayashizaki,Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2  
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159



Db 890 AAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGTACACAGTCT 949

Qy 661 AAGCTGAAGGGGAAACTGATCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTG 720

Db 950 AAACCTGAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGGTGTG 1009

Qy 721 AAATCCAGCATTGGCTGCAACACCTGGCTATTATCAGGGACCTTGGTGATGATATCTTC 780

Db 1010 AAATCCAGCATTGGCTGCAACACCTGGCTATTATCAGGGACCTTGGTGATGATATCTTC 1069

Qy 781 ACTGTACAGAGGATGAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGGATGAAACTA 840

Db 1070 ACTGTACAGAGGATGAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGGATGAAACTG 1129

Qy 841 CTCAFTGAACCTACAGCTGGTGTGGAGTGGCTGCTGTCTCAACATTTTCAAACT 900

Db 1130 CTCAFTGAACCTACAGCTGGTGTGGAGTGGCTGCTGTCTCAACATTTTCAAACT 1189

Qy 901 GTTTCCTCCAGAGTAAAGAACTTGTATTGTCTCAGTGGTGGAAATGTAGACTTAACC 960

Db 1190 GTTTCCTCCAGAGTAAAGAACTTGTATTGTCTCAGTGGTGGAAATGTAGACTTAACC 1247

Qy 961 TCCCTCCATAACTTGGGTGAAGCAGCTGAAAGGCCAGCTTCTTATCAGTCTGTTTCTGTT 1020

Db 1248 -CCTCCCTGAACCTGGGTGGGCGAGCTGAGCGCCAGCTCTTACCAGCGTTTCTGTT 1306

Qy 1021 TAA 1023

Db 1307 TAA 1309

RESULT 2

BI763753

LOCUS

DEFINITION 603047780F1 NIH\_MGC\_116 Homo sapiens cDNA clone IMAGE:5188099 5', mRNA sequence.

ACCESSION BI763753

VERSION BI763753.1 GI:15755331

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE 1 (bases 1 to 785)

JOURNAL NIH-MGC http://mgs.nci.nih.gov/.

COMMENT National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaaps-r@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: L14M11469 row: p column: 20

High quality sequence stop: 783.

Location/Qualifiers

1. 785

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5188099"

/clone\_lib="NIH\_MGC\_116"

/lab\_host="DH10B"

/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and

enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: This is a NIH\_MGC Library."

BASE COUNT 225 a 168 c 200 g 192 t

ORIGIN

Query Match 73.4%; Score 751; DB 13; Length 785;

Best Local Similarity 99.6%; Pred. No. 1.7e-214;

Matches 784; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

Qy 104 ATCAACTAACAGGGCGCAATCTTTCTTCAATGTGAATCTTCCAGAAACAGCATCTT 163

Db 1 ATCAACTAACAGGGCGCAATCTTTCTTCAATGTGAATCTTCCAGAAACAGCATCTT 60

Qy 164 -TTAAGATTCTGGTGTCTCTCAATGCCCTCAGAAGCTTGGTTCCTGTATGCTTTAGAAAG 222

Db 61 GTTAAGATTCTGGTGTCTCTCAATGCCCTCAGAAGCTTGGTTCCTGTATGCTTTAGAAAG 120

Qy 223 AAGCGGAAGCTGTTGTACTACAGCAGTGGAAACCATGGCCAGGCTCTCACTATGCT 282

Db 121 AAGCGGAAGCTG-TGTTACTACAGCAGTGGAAACCATGGCCAGGCTCTCACTATGCT 179

Qy 283 GCCAAATTGGAAGGAATTCCTGCTTATATTGTGGTGGCCCGACAGAGCTCCAGAGCTGTA 342

Db 180 GCCAAATTGGAAGGAATTCCTGCTTATATTGTGGTGGCCCGACAGAGCTCCAGAGCTGTA 239

Qy 343 AAATTTGCAATACAAAGCCTACGGAGCGTCAATTGTATCTGTGAACCTAGTATGATGCTCC 402

Db 240 AAATTTGCAATACAAAGCCTACGGAGCGTCAATTGTATCTGTGAACCTAGTATGATGCTCC 299

Qy 403 AGAGAAATTTGCAAAAGAGTTACAGAAAGAAAGAGGATCATGGTATCATCCCAAC 462

Db 300 AGAGAAATTTGCAAAAGAGTTACAGAAAGAAAGAGGATCATGGTATCATCCCAAC 359

Qy 463 CAGAGCCTGAGTATAGCTGGCAAGGCAATTTGCCCTGGAGTGTCTGAACCATGTT 522

Db 360 CAGAGCCTGAGTATAGCTGGCAAGGCAATTTGCCCTGGAGTGTCTGAACCATGTT 419

Qy 523 CCTTTGGTGGATGCACCTGCTGCTAGCTAGTGGAGGAGGAATGCTTGTGGAATAGCA 582

Db 420 CCTTTGGTGGATGCACCTGCTGCTAGCTAGTGGAGGAGGAATGCTTGTGGAATAGCA 479

Qy 583 ATTACAGTTAAGGCTCTCAAACTAGTGTGAAGGTATATGCTGTGAACCTCAATGCA 642

Db 480 ATTACAGTTAAGGCTCTCAAACTAGTGTGAAGGTATATGCTGTGAACCTCAATGCA 539

Qy 643 GATGACTCTACCACTGCAAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAA 702

Db 540 GATGACTCTACCACTGCAAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAA 599

Qy 703 ACCATAGCAGATGGTGTCAAACTCCAGCATTTGGCTTTGAACACCTGCTTATTCAGGAC 762

Db 600 ACCATAGCAGATGGTGTCAAACTCCAGCATTTGGCTTTGAACACCTGCTTATTCAGGAC 659

Qy 763 CTTGTGGATGATATCTTCACTGTACAGAGATGAATTAAGTGTGCAACCCAGCTGGTG 822

Db 660 CTTGTGGATGATATCTTCACTGTACAGAGATGAATTAAGTGTGTC-ACCAGCTGGTG 718

Qy 823 TGGGAGAGGATGAAACTACTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTG 882

Db 719 TGGGAGAGGATGAAACTACTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTG 778

Qy 883 TCTCAAC 889

Db 779 TCTCAAC 785

RESULT 3

BI763753

LOCUS

DEFINITION AU131397 NT2RP3 Homo sapiens cDNA clone NT2RP3002501 5', mRNA sequence.

ACCESSION AU131397

VERSION	AU131397.1	GI:10991751
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 848)	
AUTHORS	Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.	
TITLE	HRI human cDNA project	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3975 Fax: 81-438-52-3986 Email: genomics@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.	
FEATURES	Location/Qualifiers	
source	1..848	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="NT2RP302501"	
	/clone_lib="NT2RP3"	
	/cell_type="teratocarcinoma"	
	/cell_line="NT2"	
	/note="Vector: pME18SPL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"	
BASE COUNT	236 a 187 c 214 g 207 t	
ORIGIN		
Query Match	71.3%; Score 729.2; DB 9; Length 848;	
Best Local Similarity	98.0%; Pred. No. 6.4e-208;	
Matches	768; Conservative 0; Mismatches 12; Indels 4; Gaps 3;	
QY	1 ATGTGTGCTCAGTATTGCATCTCTTGTCTGATGTTGAAAAGCTCATCAACATTGCA 60	
DB	69 ATGTGTGCTCAGTATTGCATCTCTTGTCTGATGTTGAAAAGCTCATCAACATTGCA 128	
QY	61 GATTCTATCACCTCACACAGTGTGTAAAGCTCCATTTGAATCAACTAACAGGGCGC 120	
DB	129 GATTCTATCACCTCACACAGTGTGTAAAGCTCCATTTGAATCAACTAACAGGGCGC 188	
QY	121 AATCTTTTCTTCAATGTGAATCTTCTCCAGAAACAGGATCTTTAAGATCGTGGTCT 180	
DB	189 AATCTTTTCTTCAATGTGAATCTTCTCCAGAAACAGGATCTTTAAGATCGTGGTCT 248	
QY	181 CTCATGCGCTCAGAGCTTGTCTCTGTATGCTTTAGAAAGAACGCCGAAGCTGTGTT 240	
DB	249 CTCATGCGCTCAGAGCTTGTCTCTGTATGCTTTAGAAAGAACGCCGAAGCTGTGTT 308	
QY	241 ACTCACAGCAGTGGAAACCATCGCCAGGCTCTCACCTATGCTGCCAAATTTGAAGGAATT 300	
DB	309 ACTCACAGCAGTGGAAACCATCGCCAGGCTCTCACCTATGCTGCCAAATTTGAAGGAATT 368	
QY	301 CTGCTTTATATTGTGTGCTCCCGCCAGACAGCTCCAGACTGTAAAAAACTTGCATACAAGCC 360	
DB	369 CTGCTTTATATTGTGTGCTCCCGCCAGACAGCTCCAGACTGTAAAAAACTTGCATACAAGCC 428	
QY	361 TAGGGAGCGTCAATTGTATCTGTGAACCTAGTGATGATGATCCAGAGAAAATTTGCAAAA 420	
DB	429 TAGGGAGCGTCAATTGTATCTGTGAACCTAGTGATGATGATCCAGAGAAAATTTGCAAAA 488	
QY	421 AGAGTTTACAGAGAAACAGAGGCTCATGTGTACATCCCAACAGGAGCCCTGCCAGTGATA 480	
DB	489 AGAGTTTACAGAGAAACAGAGGCTCATGTGTACATCCCAACAGGAGCCCTGCCAGTGATA 548	
QY	481 GCTGGACAAGGGACAATTGCCCTGGAAGTGTCTGAACACAGGTTCCCTTTGGTGGATGCACGTG 540	



QY 454 CATCCCAACGAGCCTGCACTGATAGCTGGACAAGGACAAATGGCCCTGGAAGTGCTG 513  
 |||||  
 Db 422 CATCCCAACGAGCCTGCACTGATAGCTGGACAAGGACAAATGGCCCTGGAAGTGCTG 481  
 |||||  
 QY 514 AACCAAGGTTCCCTTTGGTGGATGCACTGGTGGTACCTGATAGTGGAG 559  
 |||||  
 Db 482 AACCAAGGTTCCCTTTGGTGGATGCACTGGTGGTACCTGATAGTGGAG 527  
 |||||

RESULT 6  
 BF163794 769 bp mRNA linear EST 30-OCT-2000  
 LOCUS 601769701f1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3988873 5',  
 DEFINITION mRNA sequence.

ACCESSION BF163794  
 VERSION BF163794.1 GI:11043971

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 769)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L14W9198 row: a column: 02

High quality sequence stop: 699.

Location/Qualifiers

1..769

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:3988873"

/clone\_lib="NCI\_CGAP\_Lu29"

/tissue\_type="spontaneous tumor, metastatic to mammary."

Stem cell origin."

/lab\_host="DH10B"

/note="Organ: lung; Vector: pCMV-SPORT6; Site: 1: SalI;

Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigator

providing samples: Gilbert Smith, NIH"

BASE COUNT 225 a 186 c 187 g 171 t

ORIGIN

Query Match 51.2%; Score 523.8; DB 12; Length 769;  
 Best Local Similarity 87.1%; Pred. No. 3.8e-146;  
 Matches 633; Conservative 0; Mismatches 87; Indels 7; Gaps 5;

QY 1 ATGTGCTGCTAGTATGCTATCCCTTCTGCTGATGTTGAAAAGCTCATATCAACATTGCA 60  
 |||||

Db 40 ATGTGCTGCTAGTATGCTATCCCTTCTGCTGATGTTGAAAAGCTCATATCAACATTCAA 99  
 |||||

QY 61 GATTCATATCCACCTCACACCACTGCTTAACAAGCTCCATTTTGAATCAACTAACAGGCGC 120  
 |||||

Db 100 GACTCTATCCACCTCACCCAGTGTACACAGCTCCATTTTGAATCAACTAACAGGCGC 159  
 |||||

QY 121 AATCTTTTCTCAATGTGAAGTCTTCCGAAACAGGATCTTTTAAAGATGCTGGTGCT 180  
 |||||

Db 160 AATCTTTTCTCAATGTGAAGTCTTCCGAAACAGGATCTTTTAAAGATGCTGGTGCT 219  
 |||||

QY 181 CTCATATGCGTCAGAGCTTGGTCTCTCATGC-TTTAGAAAGGAGCCCAAGCTGTTGT 239  
 |||||

Db 220 CTTAATGCCATCAGAGGCTTAATTCCTGACACGCAAGAAGAGCCCAAGCGGTAGT 279  
 |||||

QY 240 TACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCCTATGCTGCCAAATTTGGAAGGAAT 299  
 |||||

Db 280 TACTCACAGCAGCGGAAACCATGGCCAGGCTCTCACCCTATGCTGCCAAATTTGGAAGGAAT 339  
 |||||

QY 300 TCCTGCTTTATATGTGGTGGCCCGACAGCTCCAGAGCTGTAATAAATTTGCAATACAAGC 359  
 |||||

Db 340 TCCTGCTTTATATGTGGTGGCCCGACAGCTCCCAAAAGCTGCAAGAAATTTGCAATACAAGC 399  
 |||||

QY 360 CTACGAGCGTCAATTTGTTATCTGTAACCTAGTGTGATGCTCCAGAGAAATTTGTCAAA 419  
 |||||

Db 400 CTATGGAGCATCGATAGTATATCTGTGACCCAAAGTGACGAGTCCAGAGAAAGTCACTCA 459  
 |||||

QY 420 AAGAGCTTACAGAAGAAACAGAGGATCATGTTGATCATCCCAACAGGAGGCTGCAAGT 479  
 |||||

Db 460 AAGAATTTATGCAAGAAACAGAGGATCATGTTGATCATCCCAACAGGAGGCTGCAAGT 519  
 |||||

QY 480 AGCTGGACAAGGACAAATTTGGCCTGGAAGTGTGAACCCAGGTTCTTTTGGTGGATGCAC 539  
 |||||

Db 520 AGCTGGACAAGGACAAATTTGGCCTGGAAGTGTGAACCCAGGTTCTTTTGGTGGATGCAC 579  
 |||||

QY 540 GGTGTACTCTGTAGTGGAGGAGGAATGCTTGTGCTGGAATAGCAATTAAGGCTCT 599  
 |||||

Db 580 GGTGTACTCTGTAGTGGAGGAGGAATGCTTGTGCTGGAATAGCAATTAAGGCTCT 638  
 |||||

QY 600 GAAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAATGCAGATGCTGTACCAAGTC 659  
 |||||

Db 639 GAAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAATGCAGATGCTGTACCAAGTC 697  
 |||||

QY 660 CAAGCTGAAGGGAAACCTGATGCCCAATCTTTATCTCCAGAAACCATAGCATGGTGT 719  
 |||||

Db 698 TAACTGAAGGAGAACTG--GCCCAATTTAATCTCAAGAAAC--ATAGCGATGGTGT 753  
 |||||

QY 720 CAAATCC 726  
 |||||

Db 754 CAAATCC 760  
 |||||

RESULT 7

LOCUS BB609829

DEFINITION BB609829 RIKEN full-length enriched, 18 days embryo Mus musculus

ACCESSION BB609829

VERSION BB609829.1 GI:16451544

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 667)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,

, Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda

, M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,

Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakai,H., Sasaki

, D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,

Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,

Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: [genome-res@gsc.riken.go.jp](mailto:genome-res@gsc.riken.go.jp)

URL: <http://genome.gsc.riken.go.jp/>

Carninci,P., Shibata,Y., Hayatsu,M., Sugahara,Y., Shibata,K., Itoh

, M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,  
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
 Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
 ,Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa  
 ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
 Hayashizaki,Y.  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.  
 e mouse tissues.

#### FEATURES

Location/Qualifiers  
 1. 667  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_lib="RIKEN full-length enriched, 18 days embryo"  
 /dev\_stage="18 days embryo"  
 /lab\_host="SOLr"  
 /note="Site\_1: XhoI; Site\_2: SstI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN, Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5'  
 GAGAGAGAGCGCGGCACTGAGTCTTTTCTTTTCTTTT 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by  
 cap-trapper. Second strand cDNA was prepared with the  
 primer adapter of sequence [5'  
 GAGAGAGAGAGTCCAGAGCTCAATTAATTAATTAACCCCCCCCC 3'].  
 cDNA was cleaved with XhoI and SstI."  
 BASE COUNT 190 a 163 c 166 g 147 t 1 others  
 ORIGIN

Query Match 48.1%; Score 491.8; DB 10; Length 667;  
 Best Local Similarity 88.7%; Pred. No. 1.5e-136;  
 Matches 532; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
 QY 1 ATGTGTGCTCAGTATTGCTATCTCTTTGCTGATGTTGAAAAAGCTCATATCAACATTGCA 60  
 DB 68 ATGTGTGCTCAGTATTGCTATCTCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCAA 127  
 QY 61 GATTCTATCCACTCACACCACTGCTTAACAAGCTCCATTTGAATCAACTACAGGGCC 120  
 DB 128 GACTCTATCCACTCACCCAGTGTAAACAAGCTCCATTTGAATCAACTACAGGGCC 187  
 QY 121 AATCTTTTCTTCAAAATGTGAATCTTCCAGAAAAAGGATCTTTAAAGATTCTGTTGCT 180  
 DB 188 ATCTTTTCTTCAAAATGTGAGCTCTTCCAGAAATCGGTCTTTAAAGATTCTGAGTGCC 247  
 QY 181 CTCAATGCGGTGAGAGCTTGTGCTGATGCTTTAGAAAGGAGCCGAAAGCTGTGTT 240  
 DB 248 CTTAATGCCATCAGAGGCTTAATCTGTACAGCCGAGAGAGAGCCCAAGCCCGTAGTT 307  
 QY 241 ACTCACAGAGTGGAAACCATGGCCAGGCTCTCACTATGCTGCCAAATTTGAAGGAATT 300  
 DB 308 ACTCACAGAGGGAACCATGGCCAGGCTCTCACTATGCTGTAAACTTGAAGGAATT 367  
 QY 301 CTGCTTTATATGTGGTGGCCCGAGACAGCTCCAGACTGTAAAAAATCTTCAATACAAGCC 360  
 DB 368 CTGCTTTATATGTGGTGGCCCGAGACAGCTCCAGACTGTAAAAAATCTTCAATACAAGCC 427

QY 361 TAGCGAGCGCTCAATTGTTATATCTGTGAACCTAGTGTAGTCCAGAGAAAAATGTTGCAAAA 420  
 DB 428 TATGGAGCATCGATAGTACTGTGACCCCAAGTGCAGAGTCCAGAGAAAAAGGTCACCTCAA 487  
 QY 421 ACAGTTTACAGAGAAACAGAGGCATCATGTTACATCCCAACAGGAGGCTCGAGTGATA 480  
 DB 488 AGAATTATCAAGATACAGAGGCGATCTTGGTCCATCCCAACAGGAGGCTCGAGTGATA 547  
 QY 481 GCTGCAACAGGAGCAATTCCTCGAAGTCTGAACCAAGGTTCTCTTTGGTGGATGCACCTG 540  
 DB 548 GTGACAGAGGACAATTCCTCGAAGTCTGAACCAAGGTTCTCTTTGGTGGATGCACCTG 607  
 QY 541 GTGTCACCTCTAGGTGGAGGAGGAATGCTTGGTGAATAGCAATTAACAGGCTCTG 600  
 DB 608 GTGGTACCAGTAGGAGGAGGAGGAATGTTGCTGGAATAGCCATTACANTTAAGGCCCTG 667

RESULT 8  
 AV709023 652 bp mRNA linear EST 09-OCT-2000  
 LOCUS AV709023 ADC Homo sapiens cDNA clone ADCAUC06 5', mRNA sequence.  
 DEFINITION  
 AV709023  
 AV709023  
 AV709023.1 GI:10726288  
 VERSION  
 EST.  
 KEYWORDS  
 SOURCE  
 HOMO SAPIENS  
 ORGANISM  
 HOMO SAPIENS  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 AUTHORS  
 Peng,Y., Song,H., Huang,Q., Huang,Q., Gu,Y., Yang,Y., Gao,G., Xiao  
 ,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z.,  
 Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu  
 ,G., Hu,R., Chen,J., Chen,Z. and Han,Z.  
 TITLE  
 Homo sapiens cDNA ADC clones  
 JOURNAL  
 Unpublished (2000)  
 COMMENT  
 Contact: Zeguang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801919 (ex.45)  
 Fax: 86-21-50801922  
 Email: hanzge@hgc.sh.cn  
 This clone is available at CHGC in Shanghai.

FEATURES  
 Location/Qualifiers  
 1. 652  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="ADCAUC06"  
 /clone\_lib="ADC"  
 /tissue\_type="Adrenal gland"  
 /dev\_stage="Adult"  
 /lab\_host="SOLr"  
 /note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
 XhoI"  
 BASE COUNT 173 a 147 c 165 g 167 t  
 ORIGIN

Query Match 47.7%; Score 487.6; DB 10; Length 652;  
 Best Local Similarity 89.3%; Pred. No. 2.8e-135;  
 Matches 582; Conservative 0; Mismatches 14; Indels 56; Gaps 3;  
 QY 1 ATGTGTGCTCAGTATTGCTATCTCTTTGCTGATGTTGAAAAAGCTCATATCAACATTGCA 60  
 DB 56 ATGTGTGCTCAGTATTGCTATCTCTTTGCTGATGTTGAAAAAGCTCATATCAACATTGCA 115  
 QY 61 GATTCTATCCACTCACACCACTGCTTAACAAGCTCCATTTGAATCAACTACAGGGCC 120  
 DB 116 GATTCTATCCACTCACACCACTGCTTAACAAGCTCCATTTGAATCAACTACAGGGCC 175  
 QY 121 AATCTTTTCTTCAAAATGTGAATCTTCCAGAAAAAGGATCTTTAAAGATTCTGTTGCT 180  
 DB 176 AATCTTTTCTTCAAAATGTGAATCTTCCAGAAAAAGGATCTTTAAAGATTCTGTTGCT 235

QY 181 CTCATGCGCTCAGAGCTTGGTTCCTGATGCTTTAGAAAAGAGCGAAGCTGTGTT 240  
 |||||  
 Db 236 CTCATGCGCTCAGAGCTTGGTTCCTGATGCTTTAGAAAAGAGCGAAGCTGTGTT 295  
 |||||  
 QY 241 ACTCAGAGAGTGAACACATGCGCAGGCTCTCACCTATGCTGCCAATTTGGAAGAAAT 300  
 |||||  
 Db 296 ACTCAGAGAGTGAACACATGCGCAGGCTCTCACCTATGCTGCCAATTTGGAAGAAAT 355  
 |||||  
 QY 301 CCTGCTTATATTCGTGGTCCCGCAGAGCTCCAGAGCTGTAATAAATTCGAATACAGCC 360  
 |||||  
 Db 356 CCTGCTTATATTCGTGGTCCCGCAGAGCTCCAGAGCTGTAATAAATTCGAATACAGCC 415  
 |||||  
 QY 361 TACGGAGCGCTCAATTTGATATCTGTAACCTAGTAGTGAGTCCAGAGAAAATGTTCAAAA 420  
 |||||  
 Db 416 TACGGAGCGCTCAATTTGATATCTGTAACCTAGTAGTGAGTCCAGAGAAAATGTTCAAAA 472  
 |||||  
 QY 421 AGATTACAGAGAAACAGAGGCATCATGGTACATCCCAACAGGAGCGCTGCAGTGATA 480  
 |||||  
 Db 473 -----GCAGTCATA 481  
 |||||  
 QY 481 GCTGGACAGGAGCAATTCGCTGGAAGTCTGCAACAGGTTCTCTTTGGTGGATGCACTG 540  
 |||||  
 Db 482 GCTGGACAGGATCAATTCGCTGGAAGTCTGCAACAGGTTCTCTTTGGTGGATGCACTG 541  
 |||||  
 QY 541 GTGGTACCTGTAGGTGGAGGAGGAATGCTTGTGGAATAGCAATTAACAGTTAAGGCTCTG 600  
 |||||  
 Db 542 GGGGTACCTGTAAGTCGAGGATCAATGCTTGTGGAATTCGAATTCAGTTAAGGCTCT- 600  
 |||||  
 QY 601 AAACCTAGTGTAGGTATATGCTGCTGCAACCTCAATTCAGATGACTGCT 652  
 |||||  
 Db 601 AAACCTAGTGTAGGTATATGCTGCTG-ACCCTCAATTCAGATGACTGCT 651  
 |||||

## RESULT 9

BI988879  
 LOCUS 4022-49 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,  
 mRNA sequence.  
 BI988879  
 BI988879.1 GI:17959868

## ACCESSION

VERSION

## KEYWORDS

EST.

## SOURCE

house mouse.

## ORGANISM

Mus musculus.

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## AUTHORS

Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,

White, R.A., Beremand, P.D., Thomas, F.L., Gan, L. and Klein, W.H.

## TITLE

Gene expression in the developing mouse retina by EST sequencing

and microarray analysis

Nucleic Acids Res. 29 (24), 4983-4993 (2001)

21671825

Contact: Klein WH

Department of Biochemistry and Molecular Biology

University of Texas M.D. Anderson Cancer Center

Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA

Tel: 713 792 3646

Fax: 713 790 0329.

## FEATURES

Location/Qualifiers

1..600

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"

/tissue\_type="neural retina"

/note="Vector: pAMP10 (Gibco); Cloned unidirectionally."

Primer: Oligo dT. RNA Isolation: Cloned unidirectionally.

(Mannatis); Cloning Technique: CUA Cloning (Clontech).

Life Technologies); Average insert size: 1.8 Kb;

Insertion site: TACGTCACCTGAATTCGTAGTG---. Other

information regarding entire library may be found at

http://pga.swmed.edu/Data/Libraries/microarray\_cdna\_librar

ies.hlm."

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 520; Conservative

0; Mismatches 75; Indels 0; Gaps 0;

ies.hlm."

177 a 145 c 134 g 136 t 8 others

46.9%; Score 479.8; DB 13; Length 600;

87.4%; Pred. No. 5.9e-133;

0; Mismatches 75; Indels 0; Gaps 0;

ies.hlm."

1 ATCTGTGCTCAGTATTTGTCATCTCTTTGCTGATGTTGAAAAAGCTATATACACATTCGA 60

6 ATGTGTGCTCAGTACTGTCATCTCTTTGCTGATGTTGAAAAAGCTATATACACATTCGA 65

61 GATTCTATCCACCTCAGAGCTGCTTAAAGCTCCATTTTGAATCAACATTAACAGGCGC 120

66 GACTNTATCCACCTCAGAGCTGCTTAAAGCTCCATTTTGAATCAANAGCAGGCGC 125

121 AATCTTTTCTCAATTTGTAACCTCTTCCAGAAACAGGATCTTTTAAAGTTCGTGGTCT 180

126 AATCTTTTCTCAATTTGTAACCTCTTCCAGAAACAGGATCTTTTAAAGTTCGTGGTCT 185

181 CTCAATGCGCTCAGAGCTTGGTTCCTGATGCTTTAGAAAGAGCGGAAAGCTGTGTT 240

186 CTTAATGCCATCAGAGCTTAAATCTGACACCCAGAGAGAGGCGGAAAGCGTAGTT 245

241 ACTCACAGCAGTGGAAACCATGCGCAGGCTCTCACCTATGCTGCCAAATTTGGAAGAAAT 300

246 ACTCACAGCAGCGGAAACCATGCGCAGGCTCTCACCTATGCTGCCAAATTTGGAAGAAAT 305

301 CCTGCTTATATTTGCTGCGCCAGAGCTGCTTAAAGCTGTAAAGCTTGAATACAGGCC 360

306 CCTGCTTATATTTGCTGCGCCAGAGCTGCTTAAAGCTGTAAAGCTTGAATACAGGCC 365

361 TACGGAGCGCTCAATTTGTAACCTCTTCCAGAAACAGGATCTTTTAAAGTTCGTGGTCT 420

366 TATGGAGCATCGGTAGTATCTGTGACCAAGTACAGAGTCCAGAGAAAGGTCACCTCAA 425

421 AGAGTTTACAGAGAAACAGAGCATCATGGTATCATCCCAACAGGAGCGCTGCAGTGATA 480

426 AGAATTTATCAAGANACAGAGGCGCATTTGGTCCATCCCAACAGGAGCGCTGCAGTGATA 485

481 GCTGGACAGGAGCAATTTGCCCTGGAGTCTGCTGAACAGGATCTCTTTGGTGGATGACTG 540

486 GCTGGACAGGAGCAATTTGCCCTGGAGTCTGCTGAACAGGATCTCTTTGGTGGATGACTG 545

541 GTGGTACCTCTAGGTGAGGAGGAATGCTTGGTGGAAATAGCAATTAACAGTTAAGG 595

546 GTGGTACCTCTAGGTGAGGAGGAATGCTTGGTGGAAATAGCAATTAACAGTTAAGG 600

## RESULT 10

AV649313

LOCUS

AV649313 GLC Homo sapiens cDNA clone GLCBSF06 3', mRNA sequence.

ACCESSION

AV649313.1 GI:9870327

VERSION

EST.

KEYWORDS

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 697)

Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,

Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,

Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,

Hu, G., Gu, J., Chen, Z., and Han, Z.

Insight into hepatocellular carcinogenesis at transcriptome level

by comparing gene expression profiles of hepatocellular carcinoma

with those of corresponding noncancerous liver

Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

21625106

CONTACT: Zequang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China  
Tel: 86-21-50801919(ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

# FEATURES

Location/Qualifiers  
1..697  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="GLCSF06"  
/clone\_lib="GLC"  
/tissue\_type="corresponding non cancerous liver tissue"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 190 a 154 c 174 g 171 t 8 others  
ORIGIN

Query Match 46.8%; Score 479.2; DB 10; Length 697;  
Best Local Similarity 87.1%; Pred. No. 9.5e-133;  
Matches 601; Conservative 0; Mismatches 31; Indels 58; Gaps 5;

QY 1 ATGTGCTCAGTATTCATCTCTTCTGCTGATGTTGAAAAGCTCATATCAACATTGCA 60  
DB 54 ATGTGCTCAGTATTCATCTCTTCTGCTGATGTTGAAAAGCTCATATCAACATTGCA 113

QY 61 GATTCTATCCACCTCACACAGTGTAAACAGCTCCATTTTGAATCAACTAACAGGGCGC 120  
DB 114 GATCTATCCACCTCACACAGTGTAAACAGCTCCATTTTGAATCAACTAACAGGGCGC 173

QY 121 AATCTTTTCTCAATGTGAATCTTCCAGAAAACAGGATCTTTTAAAGATTCTGGTGCT 180  
DB 174 AATCTTTTCTCAATGTGAATCTTCCAGAAAACAGGATCTTTTAAAGATTCTGGTGCT 233

QY 181 CTCAATGCGCTCAGAGCTTGGTCTCTGATGCTTTAGAAAGAGCCGAAAGCTGTTGTT 240  
DB 234 CTCAATGCGCTCAGAGCTTGGTCTCTGATGCTTTAGAAAGAGCCGAAAGCTGTTGTT 293

QY 241 ACTCACAGCTGGAACCATGGCCAGGCTCTCAGCTATGCTGCGCAAAATGGAAGGAATT 300  
DB 294 ACTCACAGCTGGAACCATGGCCAGGCTCTCAGCTATGCTGCGCAAAATGGAAGGAATT 353

QY 301 CTGCTTATATGTGGTGGCCCGACAGCTCCAGACTGTAAACAACTTGCATACAAGCC 360  
DB 354 CTGCTTATATGTGGTGGCCCGACAGCTCCAGACTGTAAACAACTTGCATACAAGCC 413

QY 361 TACGAGGCTCAATGTATCTGTAACCTAGTATGATGATGATGATGATGATGATGATGAT 420  
DB 414 TACGAGGCTCAATGTATCTGTAACCTAGTATGATGATGATGATGATGATGATGATGAT 459

QY 421 AGAGTTACAGAAACAGAGGATCATGTTACATCCCAACAGGAGCCCTGCAGTGATA 480  
DB 460 -----ACCAGGAGCCCTGCAGTGATA 479

QY 481 GCTGACAGGAGCAATGCTGGAAGTGTGAACAGGTTCTCTTGGTGGATGCACTG 540  
DB 480 GCTGACAGGAGCAATGCTGGAAGTGTGAACAGGTTCTCTTGGTGGATGCACTT 539

QY 541 GTGTTACTGTAGTGGAGGAGGATGCTTCTGGAATAGCAATTACAGTTAAGGCTCTG 600  
DB 540 GTGTTACTGTAGTGGAGGAGGATGCTTCTGGAATAGCAATTACAGTTAAGGCTCTG 598

QY 601 AAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAAAATGCAGATGACTGTACCAAGTCC 660  
DB 599 AAA-CTAGTGTGAAGG-ATATGCTGTGAACCTTC-AATGCANATGACTGCTACCAAGTCC 655

QY 661 AAGCTGAGGGAGAACTGATGCCCAATCTT 690  
DB 656 AAGCTGAGGGAGAACTGATGCCCAATCTT 685

RESULT 11

BI739030

LOCUS

603359880F1 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:5367042 5',  
mRNA sequence.

ACCESSION

BI739030

VERSION

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 819)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: The Cepko Laboratory

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LAM11933 row: p column: 19

High quality sequence stop: 817.

FEATURES

source

1..819

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="IMAGE:5367042"

/clone\_lib="NIH\_MGC\_94"

/tissue\_type="retina"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 3.3 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH\_MGC Library."

BASE COUNT 236 a 208 c 192 g 183 t

ORIGIN

Query Match 44.1%; Score 451; DB 13; Length 819;

Best Local Similarity 89.0%; Pred. No. 3.1e-124;

Matches 487; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1 ATGTGCTCAGTATTGCTATCTCTTTCCTGATGTTGAAAAGCTCATATCAACATTGCA 60

DB 273 ATGTGCTCAGTATTGCTATCTCTTTCCTGATGTTGAAAAGCTCATATCAACATTGCA 332

QY 61 GATTCTATCCACCTCACACAGTGTAAACAGCTCCATTTTGAATCAACTAACAGGGCGC 120

DB 333 GACTCTATCCACCTCACCCAGTGTAAACAGCTCCATTTTGAATCAACTAACAGGGCGC 392

QY 121 AATCTTTTCTCAATGTGAATCTTCCAGAAAACAGGATCTTTTAAAGATTCTGGTGCT 180

DB 393 AATCTTTTCTCAATGTGAATCTTCCAGAAAACAGGATCTTTTAAAGATTCTGGTGCT 452

QY 181 CTCAATGCGCTCAGAGCTTGGTCTCTGATGCTTTAGAAAGAGCCGCAAGGCTGTTGT 240

DB 453 CTCAATGCGCTCAGAGCTTAAATCTGACAGCCAGAGAGAGAGAGAGAGAGAGAGAG 512

QY 241 ACTCACAGCTGGAACCATGGCCAGGCTCTCAGCTATGCTGCGCAAAATGGAAGGAATT 300

DB 513 ACTCACAGCTGGAACCATGGCCAGGCTCTCAGCTATGCTGCTTAACTGGAAGGAATT 572

QY 301 CTGCTTATATGTGGTGGCCCGACAGCTCCAGACTGTAAACAACTTGCATACAAGCC 360

DB 573 CTGCTTATATGTGGTGGCCCGACAGCTCCAGACTGTAAACAACTTGCATACAAGCC 632

QY 361 TACGAGGCTCAATGTATCTGTAACCTAGTATGATGATGATGATGATGATGATGATGAT 420

DB 633 TATGGAGCATCGATAGTATCTGTGACCCCAAGTGCAGAGTCCAGAGAGAGAGTCACTCAA 692



BASE COUNT 287 a 366 c 259 g 247 t  
ORIGIN

/lab\_host="DH10B (phage-resistant)"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
287 a 366 c 259 g 247 t

Query Match 40.1%; Score 410.4; DB 14; Length 1159;  
Best Local Similarity 91.8%; Pred. No. 66-112;  
Matches 468; Conservative 0; Mismatches 36; Indels 5; Gaps 3;  
QY 125 TTTTCTTCAATGTGAATCTTCCAGAAACAGGATCTTTTAAGATTCGTGGTCTCTCA 184  
Db 7 TTTTCTTCAATGTGAATCTTCCAGAAACAGGATCTTTTAAGATTCGTGGTCTCTCA 66  
QY 185 ATGCCGTCAGAGCTTGTCTCTGATGCTTTAGAAAGGAGCGGAAAGCTGTGTACTC 244  
Db 67 ATGCCGTCAGAGCTTGTCTCTGATGCTTTAGAAAGGAGCGGAAAGCTGTGTACTC 126  
QY 245 ACAGCAGTGGAAACCATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAATTCCTG 304  
Db 127 ACAGCAGTGGAAACCATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAATTCCTG 186  
QY 305 CTTATATTGTGTGTCGCCAGAGCTCCAGACTGTAAAAAATTCGAATACAAAGCTACG 364  
Db 187 CTTATATTGTGTGTCGCCAGAGCTCCAGACTGTAAAAAATTCGAATACAAAGCTACG 246  
QY 365 GAGCGTCAATTCTATCTACTGTGAACCTAGTATGATGATGATGATGATGATGATGATG 424  
Db 247 GAGCGTCAATTCTATCTACTGTGAACCTAGTATGATGATGATGATGATGATGATGATG 306  
QY 425 TTACAGAAAGAACAGAGCATCTGTTACATCCCAACAGGAGCTCGAGTGATAGCTG 484  
Db 307 TTACAGAAAGAACAGAGCATCTGTTACATCCCAACAGGAGCTCGAGTGATAGCTG 366  
QY 485 G-ACAAAGGACAAATTGCCCTTGAAGTGTGAACAGGTTCTCTTGGTGGATGCA-CTGGT 542  
Db 367 GCACAAGGGACAAATTGCCCTTGAAGTGTGAACAGGTTCTCTTGGTGGATGCACTGT 426  
QY 543 GTACCTCTAGTGTGAGGAGGAATCTTCTGCTG----AATGACATTTACAGTTAGGCTC 598  
Db 427 GTACCTCTAGTGTGAGGAGGAATCTTCTGCTGGAATAAGCCCTCTACAGTTAGGCTC 486  
QY 599 TGAACACCTAGTGTGAAGTATATGCTGCTG 628  
Db 487 CTGACACCCCGCTGAAAGTATTATGCTG 516

RESULT 14  
BQ885621  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BQ885621  
AGENCY: 8754012 NIH\_MGC\_130 Mus musculus cdna clone IMAGE:6332599  
5' mRNA sequence.  
BQ885621  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 918)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,  
Ph.D.  
cdna Library Preparation: ResGen, Invitrogen Corp  
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM13790 row: h column: 08  
High quality sequence stop: 609.  
Location/Qualifiers  
1..918  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6332599"  
/clone\_lib="NIH\_MGC\_130"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.95 kb.  
Constructed by ResGen, Invitrogen Corp. Note: this is a  
NIH\_MGC Library."  
247 a 230 c 230 g 203 t 8 others

BASE COUNT 247 a 230 c 230 g 203 t 8 others  
ORIGIN

Query Match 39.3%; Score 402.4; DB 14; Length 918;  
Best Local Similarity 83.7%; Pred. No. 1.4e-109;  
Matches 512; Conservative 0; Mismatches 92; Indels 8; Gaps 5;  
QY 1 ATGTGCTGCTAGTATGCTATCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCGA 60  
Db 308 ATGTGCTGCTAGTATGCTATCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCGA 367  
QY 61 GATTCTATCCACCTCACACCAGTGTACAAAGCTCCATTTTGAATCAACATCAAGGCGC 120  
Db 368 GACTCTATCCACCTCACACCAGTGTACAAAGCTCCATTTTGAATCAACATCAAGGCGC 427  
QY 121 AATCTTTTCTTCAAAATGTGAACCTCTTCCAGAAACAGGATCTTTTAAAGATTCGTGCTG 180  
Db 428 AATCTTTTCTTCAAAATGTGAACCTCTTCCAGAAACAGGATCTTTTAAAGATTCGTGCTG 487  
QY 181 CTCAATGCGCTGAGAGCTTGTGCTGATGCTTTAGAAAGGAGCCGAAAGCTGTGTT 240  
Db 488 CTCAATGCGCTGAGAGCTTGTGCTGATGCTTTAGAAAGGAGCCGAAAGCTGTGTT 547  
QY 241 ACTCACACAGTGGAAACCATGCGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAT 300  
Db 548 ACTCACACAGTGGAAACCATGCGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAT 607  
QY 301 CTGCTCTATATTTGCTGCCAGACAGCTCCAGAGCTGTAAACAACTTCAATACAAAGC 360  
Db 608 CTGCTCTATATTTGCTGCCAGACAGCTCCAGAGCTGTAAACAACTTCAATACAAAGC 667  
QY 361 TACGAGGCTCAATTTGATATCTGTGAACCTAGTATGATGATGATGATGATGATGATGATG 420  
Db 668 TATGAGGCTCAATTTGATATCTGTGAACCTAGTATGATGATGATGATGATGATGATGATG 726  
QY 421 AGAGTTACAGAGAAACAGAGGATCATGCTATCCCAACAGGAGCTCCAGTGTGATA 480  
Db 727 AGAGTTACAGAGAAACAGAGGATCATGCTATCCCAACAGGAGCTCCAGTGTGATA 786  
QY 481 GCTGCAAGGAGCAATTTGCCCTTGAAGTGTGAACAGGTTTCCCTTTGG-TGGTCACT 539  
Db 787 GCTGCAAGGAGCAATTTGCCCTTGAAGTGTGAACAGGTTTCCCTTTGGTGTAGTCACT 846  
QY 540 -GGTGGTACCTCTAGGT--GGAGGAGGAATGCTGTGTTGAATA--GCAATTACAGTTAA 593  
Db 847 GGTGGTACCTCTAGGT--GGAGGAGGAATGCTGTGTTGAATA--GCAATTACAGTTAA 906  
QY 594 GGTCTGGAAC 605  
Db 907 GGTCTGGAAC 918

RESULT 15  
B1738966  
LOCUS  
DEFINITION

B1738966  
603359780F1 NIH\_MGC\_94 Mus musculus cdna clone IMAGE:5367043 5',  
linear EST:20-SEP-2001

DB 694 CAAAGATTAT-TGCAGAAACAGAGGCGATCTTTGGTCCATCCCAACCAGAGCCTGCAGTG 752  
QY 478 ATAGCTGGACAGGAGCAATTCGCCCTGGAAAGTGTGAACCAAGGTTCCCTTTG 528  
DB 753 ATAGCTGGACAGGAGCAATTCGCCCTGGAAAGTGTGAACCAAGGTTCCCTTTG 803

Search completed: June 24, 2003, 10:54:14  
Job time : 1468 secs

Accession BI738966  
Version BI738966.1 GI:15715979  
Keywords EST.  
Source house mouse.  
Organism Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 823)  
NIH-MGC <http://mgi.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM11933 row: p column: 20  
High quality sequence stop: 799.  
Location/Qualifiers  
1..823  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5367043"  
/clone\_lib="NIH\_MGC\_94"  
/tissue\_type="retina"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 3.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
237 a 208 c 193 g 184 t 1 others

#### FEATURES

source

Query Match 38.3%; Score 391.8; DB 13; Length 823;  
Best Local Similarity 88.5%; Pred. No. 2e-106;  
Matches 470; Conservative 0; Mismatches 57; Indels 4; Gaps 4;  
QY 1 ATGTGTCTCAGTATGTGATCTCTTTGCTGATGTTGAAAGCTCATATCAACATTGCA 60  
DB 274 ATGTGTCTCAGTATGTGATCTCTTTGCTGATGTTGAAAGCTCATATCAACATTGCA 333  
QY 61 GATCTATCCACCTCACACAGTCTTAACAGCTCCATTTGAATCAACTACAGGGCGC 120  
DB 334 GACTCTATCCACCTCACACAGTCTTAACAGCTCCATTTGAATCAACTACAGGGCGC 393  
QY 121 AATCTTTTCTCAATGTGAATCTTCCAGAAACAGGATCTTTTAAGATTCTGGTGTCT 180  
DB 394 AATCTTTTCTCAATGTGAATCTTCCAGAAACAGGATCTTTTAAGATTCTGGTGTCT 453  
QY 181 CTAAATGCGGTGAGAGTGTGGTCTCTGATGTTAGAAAGGAGCCGAAAGCTGTTGTT 240  
DB 454 CTAAATGCGGTGAGAGTGTGGTCTCTGATGTTAGAAAGGAGCCGAAAGCTGTTGTT 513  
QY 241 ACTCACAGCAGTGGAAACCATG-GCCAGGCTCTACCTATGCTGCCAATTTGGAGGAAT 299  
DB 514 ACTCACAGCAGTGGAAACCATGNGCGCAAGCTCTACCTATGCTGCCAATTTGGAGGAAT 573  
QY 300 TCCTGCTTATATTTGGTGGCCCGCCAGACAGTCCAGACTGTAA-AAACTTGCATATCAAG 358  
DB 574 TCCTGCTTATATTTGGTGGCCCGCCAGACAGTCCAGACTGTAA-AAACTTGCATATCAAG 633  
QY 359 CCTACGAGCGTCAATTTGTATAC-TGTGAACCTAGTGTGATGATCCAGAGAAATGTTGCA 417  
DB 634 CCTACGAGCGTCAATTTGTATAC-TGTGAACCTAGTGTGATGATCCAGAGAAATGTTGCA 693  
QY 418 AAAAGAGTTACAGAGAAACAGAGGATCATGTTACATCCCAACCAGGACCTCCAGTG 477  
||||| ||| | |||||||||||||| ||| |||||||||||||| ||| |||||||||||||| ||| |||

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 06:16:55 ; Search time 77 Seconds  
(without alignments)  
4074.420 Million cell updates/sec

Title: US-09-889-609B-9

Perfect score: 1023

Sequence: 1 atgtgtgtcagattgcat.....atcagctgtttctgttttaa 1023

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2.6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1023	100.0	1770	4	US-09-789-300A-1
2	1020	99.7	1020	4	US-09-789-300A-3
c 3	62.4	6.1	2299	4	US-09-153-599A-1
4	57	5.6	1545	2	US-08-628-039-1
5	57	5.6	1545	2	US-08-628-039-5
6	57	5.6	1545	2	US-08-628-039-7
7	57	5.6	1545	2	US-08-628-039-8
8	57	5.6	1545	2	US-08-673-388-1
9	57	5.6	1545	2	US-08-673-388-5
10	57	5.6	1545	2	US-08-673-388-7
11	57	5.6	1545	2	US-08-673-388-8
12	57	5.6	1545	2	US-08-614-877-1
13	57	5.6	1545	2	US-08-614-877-5
14	57	5.6	1545	2	US-08-614-877-7
15	57	5.6	1545	2	US-08-614-877-8
16	57	5.6	1545	3	US-08-912-205-1
17	57	5.6	1545	3	US-08-912-205-5
18	57	5.6	1545	3	US-08-912-205-7
19	57	5.6	1545	3	US-08-912-205-8
20	57	5.6	1545	4	US-09-440-400-1
21	57	5.6	1545	4	US-09-440-400-5
22	57	5.6	1545	4	US-09-440-400-7
23	57	5.6	1545	4	US-09-440-400-8
24	49.2	4.8	12720	1	US-08-403-866-11
c 25	46.8	4.6	7218	1	US-08-232-463-14
c 26	46	4.5	2698	4	US-08-936-165A-207
27	38.4	3.8	1485	4	US-09-088-435-2

28	36.6	3.6	6211	4	US-08-961-527-8	Sequence 8, Appli
29	36	3.5	1275	4	US-09-134-001C-1039	Sequence 1039, Ap
30	34.2	3.3	2085	2	US-08-283-917-8	Sequence 8, Appli
31	34.2	3.3	2085	2	US-08-961-716-8	Sequence 8, Appli
32	33.8	3.3	1547	4	US-09-356-818A-1	Sequence 1, Appli
33	33.8	3.3	1910	1	US-08-247-902A-1	Sequence 1, Appli
34	33.8	3.3	1910	5	PCT-US93-10541-1	Sequence 1, Appli
35	33.8	3.3	3257	5	PCT-US91-09784-1	Sequence 1, Appli
36	33.4	3.3	1274	4	US-09-227-357-75	Sequence 75, Appli
37	32.6	3.2	8494	4	US-08-961-527-163	Sequence 163, App
38	32	3.1	1557	4	US-09-134-001C-1131	Sequence 1131, Ap
39	32	3.1	3319	2	US-08-960-022-19	Sequence 19, Appli
40	32	3.1	4651	2	US-08-449-644-4	Sequence 4, Appli
41	32	3.1	4651	2	US-08-087-244A-4	Sequence 4, Appli
42	31.8	3.1	351	4	US-09-556-877-279	Sequence 279, App
43	31.8	3.1	351	4	US-09-620-412C-279	Sequence 279, App
44	31.6	3.1	289	4	US-09-007-005-17	Sequence 17, Appli
45	31.6	3.1	289	4	US-09-244-796-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1  
US-09-789-300A-1  
; Sequence 1, Application US/09789300A  
; Patent No. 6458576  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; APPLICANT: Rudolph-Owen, Laura A.  
; TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate  
; TITLE OF INVENTION: Dependent Enzyme Family Member and Uses Therefore  
; FILE REFERENCE: 35800/208926  
; CURRENT APPLICATION NUMBER: US/09/789,300A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: US 60/183,208  
; PRIOR FILING DATE: 2000-02-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1770  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (69)...(1091)  
US-09-789-300A-1

Query Match	100.0%	Score 1023;	DB 4;	Length 1770;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1023;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
QY	1	ATGTGTGTCAGTATTGTCATCTCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCGA	60	
DB	69	ATGTGTGTCAGTATTGTCATCTCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCGA	128	
QY	61	GATTCCTATCCACTCACACAGTCTCAACAGCTCCATTTTGATCAACTTAACAGGCGC	120	
DB	129	GATTCCTATCCACTCACACAGTCTCAACAGCTCCATTTTGATCAACTTAACAGGCGC	188	
QY	121	AATCTTTTCTTCAAAATGTGAAGTCTTCCAGAAAAAGAGGATCTTTTAAGATTCGTGCTGT	180	
DB	189	AATCTTTTCTTCAAAATGTGAAGTCTTCCAGAAAAAGAGGATCTTTTAAGATTCGTGCTGT	248	
QY	181	CTCAATGCGCTCAGAAAGCTTGTTCTGATGCTTTTGAAGAGGAGCGGAAAGCTGTGTT	240	
DB	249	CTCAATGCGCTCAGAAAGCTTGTTCTGATGCTTTTGAAGAGGAGCGGAAAGCTGTGTT	308	
QY	241	ACTCACAGAGTGGAAACCATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAGAAAT	300	
DB	309	ACTCACAGAGTGGAAACCATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAGAAAT	368	
QY	301	CCTGCTTATATTGTGTGTCGCCCCAGACAGCTCCAGAGCTGTAAAAAACTTGCATACAGGC	360	

Db 369 CTTCTTATATGTGGTGGCCACAGAGCTCCAGACTGTAAAAAATTTGCAATACAAGCC 428  
Qy 361 TACGGAGCGCAATTTGTATCTGTGAACCTAGTATGATGAGTCCAGAGAAAATTTGCAAAA 420  
Db 429 TACGGAGCGCAATTTGTATCTGTGAACCTAGTATGATGAGTCCAGAGAAAATTTGCAAAA 488  
Qy 421 AGAGTTACAGAAGAAACAGAGGATCATGGTACATCCCAACAGGAGCGTGCAGTGATA 480  
Db 489 AGAGTTACAGAAGAAACAGAGGATCATGGTACATCCCAACAGGAGCGTGCAGTGATA 548  
Qy 481 GCTGGACAAGGACAATTTGCCCTGGAGTGTCTGAACCAAGTTCCTTTGGTGGATGCACTG 540  
Db 549 GCTGGACAAGGACAATTTGCCCTGGAGTGTCTGAACCAAGTTCCTTTGGTGGATGCACTG 608  
Qy 541 GTGTACTCTGTAGTGGAGGAGGAATGCTTGTGGATAGCAATTCAGTTAAGGCTCTG 600  
Db 609 GTGTACTCTGTAGTGGAGGAGGAATGCTTGTGGATAGCAATTCAGTTAAGGCTCTG 668  
Qy 601 AAACCTAGTGTGAAGGTATATGCTGTGAACCTCAAATGCAGATGACTGCTACCAAGTCC 660  
Db 669 AAACCTAGTGTGAAGGTATATGCTGTGAACCTCAAATGCAGATGACTGCTACCAAGTCC 728  
Qy 661 AAGCTGAAGGGAAACCTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTC 720  
Db 729 AAGCTGAAGGGAAACCTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTC 788  
Qy 721 AAATCCAGCATTTGGCTTGAACACCTGGCTTATATCAGGACCTTTGTGGATGATATCTTC 780  
Db 789 AAATCCAGCATTTGGCTTGAACACCTGGCTTATATCAGGACCTTTGTGGATGATATCTTC 848  
Qy 781 ACTGTACAGAGGATGAATTAAGTGTGAACCCAGCTGTGTGGAGAGGATGAACACTA 840  
Db 849 ACTGTACAGAGGATGAATTAAGTGTGAACCCAGCTGTGTGGAGAGGATGAACACTA 908  
Qy 841 CTCATTGAACCTACAGCTGTGTGGAGTGGCTGTGTCTCAACATTTTCAAACT 900  
Db 909 CTCATTGAACCTACAGCTGTGTGGAGTGGCTGTGTCTCAACATTTTCAAACT 968  
Qy 901 GTTTCCTCCAGAGTAAGAACATTTGTATTGTCTCAGTGTGGAAATGTAGACTTAAACC 960  
Db 969 GTTTCCTCCAGAGTAAGAACATTTGTATTGTCTCAGTGTGGAAATGTAGACTTAAACC 1028  
Qy 961 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTTCTGTT 1020  
Db 1029 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTTCTGTT 1088  
Qy 1021 TAA 1023  
Db 1089 TAA 1091

## RESULT 2

US-09-789-300A-3  
; Sequence 3, Application US/09789300A  
; Patent No. 6458576  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; APPLICANT: Rudolph-Owen, Laura A.  
; TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate  
; TITLE OF INVENTION: Dependent Enzyme Family Member and Uses Therefore  
; FILE REFERENCE: 35800/208926  
; CURRENT APPLICATION NUMBER: US/09/789,300A  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: US 60/183,208  
; PRIOR FILING DATE: 2000-02-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1020  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-789-300A-3

Query Match 99.7%; Score 1020; DB 4; Length 1020;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTGTCTCAGTATTCATCTCTCTTTGGCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
Db 1 ATGTGTCTCAGTATTCATCTCTCTTTGGCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
Qy 61 GATTCTATCCACCTCACACCAGTCTTAACAAGCTCCATTTTGAATCAACTAACAGGGCG 120  
Db 61 GATTCTATCCACCTCACACCAGTCTTAACAAGCTCCATTTTGAATCAACTAACAGGGCG 120  
Qy 121 AATCTTTTCTTCAAAATGTGAACCTCTCCAGAAAACAGGATCTTTTAAAGATTCGTGGTCT 180  
Db 121 AATCTTTTCTTCAAAATGTGAACCTCTCCAGAAAACAGGATCTTTTAAAGATTCGTGGTCT 180  
Qy 181 CTCATGCGCTCAGAAAGCTTTGGTTCCTGATGCTTTAGAAAAGAGCCGAAAGCTGTTGTT 240  
Db 181 CTCATGCGCTCAGAAAGCTTTGGTTCCTGATGCTTTAGAAAAGAGCCGAAAGCTGTTGTT 240  
Qy 241 ACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGGAGGAAT 300  
Db 241 ACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGGAGGAAT 300  
Qy 301 CCTGCTTATATTTGTGTGCCCCAGACAGCTCCAGACTGTAAAAAATTTGCAATACAAGCC 360  
Db 301 CCTGCTTATATTTGTGTGCCCCAGACAGCTCCAGACTGTAAAAAATTTGCAATACAAGCC 360  
Qy 361 TAGCGAGCGTCAATTTGTATCTGTGAACCTAGTATGAGTCCAGAGAAAATTTGCAAAA 420  
Db 361 TAGCGAGCGTCAATTTGTATCTGTGAACCTAGTATGAGTCCAGAGAAAATTTGCAAAA 420  
Qy 421 AGAGTTACAGAAACAGAGGATCATGGTACATCCCAACAGGAGCTGCAGTGATA 480  
Db 421 AGAGTTACAGAAACAGAGGATCATGGTACATCCCAACAGGAGCTGCAGTGATA 480  
Qy 481 GCTGGACAAGGACAATTTGCCCTGGAGTGTCTGAACCAAGCTTCTTTGGTGGATGCACTG 540  
Db 481 GCTGGACAAGGACAATTTGCCCTGGAGTGTCTGAACCAAGCTTCTTTGGTGGATGCACTG 540  
Qy 541 GTGTACCTGTAGTGGAGGAGGAATGCTTTGGTGAATAGCAATTAAGTTAAGGCTCTG 600  
Db 541 GTGTACCTGTAGTGGAGGAGGAATGCTTTGGTGAATAGCAATTAAGTTAAGGCTCTG 600  
Qy 601 AAACCTAGTGTGAAGTATATGCTGTAACCTCAAATGCAGATGACTGCTACCAGTCC 660  
Db 601 AAACCTAGTGTGAAGTATATGCTGTAACCTCAAATGCAGATGACTGCTACCAGTCC 660  
Qy 661 AAGCTGAAGGGAAACCTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCT 720  
Db 661 AAGCTGAAGGGAAACCTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCT 720  
Qy 721 AAATCCAGCATTTGGCTTGAACACCTGGCTTATATCAGGAGCTTGTGGATGATATCTTC 780  
Db 721 AAATCCAGCATTTGGCTTGAACACCTGGCTTATATCAGGAGCTTGTGGATGATATCTTC 780  
Qy 781 ACTGTACAGAGGATGAATTAAGTGTGAACCCAGCTGTGTGGAGAGGATGAACACTA 840  
Db 781 ACTGTACAGAGGATGAATTAAGTGTGAACCCAGCTGTGTGGAGAGGATGAACACTA 840  
Qy 841 CTCATTGAACCTACAGCTGTGTGGAGTGGCTGTGTCTCAACATTTTCAAACT 900  
Db 841 CTCATTGAACCTACAGCTGTGTGGAGTGGCTGTGTCTCAACATTTTCAAACT 900  
Qy 901 GTTTCCTCCAGAGTAAGAACATTTGTATGTGCTCAGTGGTGGAAATGTAGACTTAAACC 960  
Db 901 GTTTCCTCCAGAGTAAGAACATTTGTATTTGTGCTCAGTGGTGGAAATGTAGACTTAAACC 960  
Qy 961 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTTCTGTT 1020  
Db 961 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTTCTGTT 1020

## RESULT 3

US-09-153-599A-1/c  
: Sequence 1, Application US/09153599A  
: Patent No. 6420177  
: GENERAL INFORMATION:  
: APPLICANT: Weber, J. Mark  
: APPLICANT: Luu, B. Minh  
: TITLE OF INVENTION: Method for Strain Improvement of  
: Bacterium Producing Erythromycin  
: NUMBER OF SEQUENCES: 13  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Rockey, Milnamow & Katz, Ltd.  
: STREET: 180 N. Stetson Avenue, 2 Prudential Plaza  
: CITY: Chicago  
: STATE: Illinois  
: COUNTRY: U.S.A.  
: ZIP: 60601  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: FILING DATE: US/09/153,599A  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Mueller, Lisa V.  
: REGISTRATION NUMBER: 38,978  
: REFERENCE/DOCKET NUMBER: FER2159P0041US  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 312-616-5400  
: TELEFAX: 312-616-5460  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 2299 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: DNA (genomic)  
US-09-153-599A-1

Query Match 6.1%; Score 62.4; DB 4; Length 2299;  
Best Local Similarity 49.5%; Pred. No. 4.1e-10;  
Matches 190; Conservative 0; Mismatches 191; Indels 3; Gaps 1;

Qy	509	TGCTGAACCAAGTTCTTTGGTGATGACACTGGTGTTACTGTAGTGGAGGAGGAATGC	568
Db	2296	TCCTGGAGCAGTTGCCGAGCTCCGGACAGTGTGTGCTCCCGAGGCGCGGGCTGG	2237
Qy	569	TTGCTGGAATAGCAATTACAGTTAAGGCTCTGAACCTAGTGTCAAGGTATATGCTGCTG	628
Db	2236	TCAGCGGCATCGCCGCGCGCTGAAGCGGAGCACCAGGTCGCGGTCTCGCCGCTC	2177
Qy	629	AACCTCAATGACGATGACTGTACAGTCCAGCTGAAGGGGAAACTGATGCCCAATC	688
Db	2176	AAGCGGAGCAGCGCCGCTGCGCCGCTGCTGGCGGCGGAAACCGTTGCTCTGC	2117
Qy	689	TTTATCTCCAGAAACCATAGCAGATGGTGTCAAAATCCAGCATTTG---GCTTGAACACCT	745
Db	2116	TCGACACCCAGCGGACGATGGCGACGCGATCGCGGTCCCGCGCGAGCGAGCTGACCT	2057
Qy	746	GGCTATTATCAGGACCTTGTGGATCATATCTTCACTGTACAGAGGATGAATTAAGT	805
Db	2056	TCGCCCCAGCTCAGCAGCTCGTACAGCATGCTCAGGGTGGGAGGAGGCGCTCTCGC	1997
Qy	806	GTGCAACCCAGCTGGTGGGAGGATGAACCTACTCATTTGAACCTACAGCTGGTGTG	865
Db	1996	GCGGCTGCTGTGCTGCTGGAAGCGGAGCTGGTGGTCGAACCCCGCGGCTCGCGG	1937
Qy	866	GAGTGGCTGCTGTGCTCTCAAC	889

Db 1936 CGTCCGCCGAGCTGCTGGAGCAC 1913

## RESULT 4

US-08-628-039-1  
: Sequence 1, Application US/08628039  
: Patent No. 5942660  
: GENERAL INFORMATION:  
: APPLICANT: Gruys, Kenneth J.  
: APPLICANT: Mitsky, Timothy A.  
: APPLICANT: Kishore, Ganesh M.  
: APPLICANT: Slater, Steven C.  
: APPLICANT: Padgett, Stephen R.  
: APPLICANT: Stark, David M.  
: APPLICANT: Hinchey, Maud A. W.  
: APPLICANT: Clemente, Thomas E.  
: APPLICANT: Connor-Ward, Danielle V.  
: APPLICANT: Fedele, Mary J.  
: APPLICANT: Fry, Joyce E.  
: APPLICANT: Howe, Arlene R.  
: APPLICANT: Rozman, Renee J.  
: TITLE OF INVENTION: Methods of Optimizing Substrate Pools and  
: Biosynthesis of poly-B-hydroxybutyrate-co-poly-B-hydroxyval  
: TITLE OF INVENTION: in Bacteria and Plants  
: NUMBER OF SEQUENCES: 11  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Gary M. Bond, Monsanto Company, A3SB  
: STREET: 800 No. 5942660th Lindbergh Boulevard  
: CITY: St. Louis  
: STATE: Missouri  
: COUNTRY: USA  
: ZIP: 63167  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: FILING DATE: US/08/628,039  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER:  
: FILING DATE: 13-MAR-1996  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Bond, Gary  
: REGISTRATION NUMBER: 29,283  
: REFERENCE/DOCKET NUMBER: 38-21(13585)A  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (314)694-3412  
: TELEFAX: (314)695-5435  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1545 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: linear  
: MOLECULE TYPE: DNA (genomic)  
US-08-628-039-1

Query Match 5.6%; Score 57; DB 2; Length 1545;  
Best Local Similarity 43.6%; Pred. No. 1.9e-08;  
Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Qy	216	AGAAAGGAGCCGAAAGCTGTTGTTACTACAGCAGTGGAAACCATGGCCAGGCTCTCAC	275
Db	225	AGAACAGAGGCGCACGCGTGATCACTGCTTCGCGGTAAACACCGCGAGGCGCTCGC	284
Qy	276	CTATGCTGCCAAATTTGAAGAATTCCTGCTTATATTGTTGGTGGCCCGACAGCTCCAGA	335
Db	285	GTTTCTTCTGCGCGGTAGGCGGTGAAGCGCTGATGCTTATGCAACCGCCACCGCGGA	344
Qy	336	CTGTAAAAAACTTGCAATACAAGCCTACGGAGCGCTCAATTGTATACTGTGAACCTAGTGA	395

Db 345 CATCAAGTCAGCGGCTGCGCGCTTCGGCGGCAAGTGTCTCCAGCGCGCAACTT 404  
Qy 396 TGAGTCCAGAGAAATGTTGCAAAAAGAGTTTACAGAAGAAACAGAGGCATCATGGTACA 455  
Db 405 TGATGAAGCGAAACGCAAGCGATCGCAACTGTACACAGCAGGCGTTACCTGGGTGCC 464  
Qy 456 TCCCAACACGAGGCGCTCAGTGTAGCTGCAAGAGGACAAATTCCTTGGAAAGTGTCAA 515  
Db 465 GCCGTTCCAGCATCCGATGTTGATTGCGGCAAGGACGCTGCGCTGGAAGTGTCCA 524  
Qy 516 CCAGGTTCTTTGTGTGATGCACCTGGTGTACCTGTAGTGGAGGAGGAATGCTTGTGG 575  
Db 525 GCAGGAGCGCCCATCGACCGCGTATTGTCAGAGTGGCGGCGCGTCTGGCTGTG 584  
Qy 576 AATGACAAATACAGTTAAGGCTGTAAGCTAGTGTGAAGCTATATGCTGTGAACCTC 635  
Db 585 CGTGGCGGTGTGATCAAACTGATGCGCAATCAAGTGTACCGGTGATCGCGCGGT 644  
Qy 636 AATGACAGATGACTGTACCAAGCTCAAGCTGAAGGGAAAGTGTGCCCCAATCTTTATCC 695  
Db 645 AGACTCGCGCTGCTGAAAGCAGCGCTGGATGCGGGTCAATCGGTGATCTGCGCGCGT 704  
Qy 696 TCCAGAAACCATACGATGGTCAAAATCCAGCATTTGGCTTGAACACCTGGCCCTATAT 755  
Db 705 AGGCTATTGCTGAAGCGGTAGCGGTAAACGATCGGTGACGAAACCTTCCGTTTATG 764  
Qy 756 CAGGACCTGTGTGATGATCTTCACTGTGCACAGAGGATGAAT 800  
Db 765 CCAGAGTATCTCAGCAGATCATCACCGTCGATAGCGATGCGAT 809

RESULT 5  
US-08-628-039-5  
; Sequence 5, Application US/08628039  
; Patent No. 5942660  
; GENERAL INFORMATION:  
; APPLICANT: Gruys, Kenneth J.  
; APPLICANT: Mitsky, Timothy A.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stark, David M.  
; APPLICANT: Hinchee, Maud A. W.  
; APPLICANT: Connor-Ward, Dannelle V.  
; APPLICANT: Fedele, Mary J.  
; APPLICANT: Fry, Joyce E.  
; APPLICANT: Howe, Arlene R.  
; APPLICANT: Rozman, Renee J.  
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and  
; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalerate  
; TITLE OF INVENTION: in Bacteria and Plants  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gary M. Bond, Monsanto Company, A3SB  
; STREET: 800 No. 5942660th Lindbergh Boulevard  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63167  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/628,039  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE: 13-MAR-1996

ATTORNEY/AGENT INFORMATION:  
; NAME: Bond, Gary  
; REGISTRATION NUMBER: 29,283  
; REFERENCE/DOCKET NUMBER: 38-21(13585)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314)694-3412  
; TELEFAX: (314)695-5435  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1545 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-628-039-5

Query Match 5.6%; Score 57; DB 2; Length 1545;  
Best Local Similarity 43.6%; Pred. No. 1.9e-08;  
Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Qy 216 AGAAGCAAGCCGAAAGCTGTTCTTACTCAGCAGTGGAAACCATGCCAGGCTCTCAC 275  
Db 225 AGAACAGAAAGCCAGCGGTGATCACTGCTTTCGGGGTAACCAACGCGGCGTGC 284  
Qy 276 CTATGCTGCCAAATTTGAAGAAATTCCTGCTTATATTGTTGTCGCCACAGCAGCTCCAGA 335  
Db 285 GTTTCTCTTCGCGGTTAGCGGTGAAGGCCCTGATCGTTATGCCAACGCCACCGCGGA 344  
Qy 336 CTGTAAAAAATTTGCAATACAGCCTACGAGCGTCAATTTGTTATCTGTGAACCTAGTGA 395  
Db 345 CATCAAAAGTCGACCGGCTTCGGCGGGAAGTGTCTCCACGCGCGCAACTT 404  
Qy 396 TGAGTCCAGAGAAATTTGCAAAAGAGTTACAGAGAAACAGAGGCATCATGGTACA 455  
Db 405 TGATGAAGCAAGCGCAAGCGATCGAACTGTTCACAGCAGCGGGTTACCTGGGTGCC 464  
Qy 456 TCCCAACAGGAGCGCTCAGTGTAGCTGGACAAGGAGCAATTCGCCCTGGAAGTGTCAA 515  
Db 465 GCCGTTCCGACCATCCGATGTGATTGCGGGCAAGGACGCTGGCGCTGGAAGTGTCCA 524  
Qy 516 CCAGGTTCTTTGTGTGATGCATCTGAGTGTACCTGTAGTGTAGTGGAGGAGGAATGTTGTGG 575  
Db 525 GCAGGACGCGCCATCTCGACCGCGTATTGTTGCCAGTTCGGCGGCGGCTGTGGCTGTG 584  
Qy 576 AATGACAAATACAGTTAAGGCTCTGAAACCTAGTGTGAAGGTATATCTCTGAACCTC 635  
Db 585 CGTGGCGGTCTGATCAACAACTGATGCGCGCAAAATCAAAGTATCCCGCTAGAAGCGGA 644  
Qy 636 AATGACAGATGACTGTCTTACCAGTCCAGCTGAAGGGGAAACTGATGCCCAATCTTTATCC 695  
Db 645 AGACTCCGCGCTGCTGAAAGCAGCGCTGGATGCGGGTCACTCCGTTGATCTGCGCGCGT 704  
Qy 696 TCCAGAAACCATACGAGATGGTGTCAAAATCCAGATTTGGCTTGAACACCTGGGCTATAT 755  
Db 705 AGGCGTATTGCTGAAGCGGTAGCGGTAAACGATCGGTGACGAAACCTTCCGTTTATG 764  
Qy 756 CAGGAGCCTTGTGATGATATCTTCACTGTGCACAGAGGATGAAT 800  
Db 765 CCAGAGTATCTCAGCAGCATCATCACCGTCGATAGCGATGCGAT 809

RESULT 6  
US-08-628-039-7  
; Sequence 7, Application US/08628039  
; Patent No. 5942660  
; GENERAL INFORMATION:  
; APPLICANT: Gruys, Kenneth J.  
; APPLICANT: Mitsky, Timothy A.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stark, David M.  
; APPLICANT: Hinchee, Maud A. W.

APPLICANT: Clemente, Thomas E.  
 APPLICANT: Connor-Ward, Dannette V.  
 APPLICANT: Fedele, Mary J.  
 APPLICANT: Fry, Joyce E.  
 APPLICANT: Howe, Arlene R.  
 APPLICANT: Rozman, Renee J.  
 TITLE OF INVENTION: Methods of Optimizing Substrate Pools and  
 TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalerate  
 TITLE OF INVENTION: in Bacteria and Plants  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Gary M. Bond, Monsanto Company, A35B  
 STREET: 800 No. 5942660th Lindbergh Boulevard  
 CITY: St. Louis  
 STATE: Missouri  
 COUNTRY: USA  
 ZIP: 63167  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/628,039  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE: 13-MAR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bond, Gary  
 REGISTRATION NUMBER: 29,283  
 REFERENCE/DOCKET NUMBER: 38-21(13585)A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (314)694-3412  
 TELEFAX: (314)695-5435  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1545 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 PS-08-628-039-7

[illegible]

QY 576 AATAGCAATTACAGTTAAGGCTCTGAACCTAGTGTGAAGGTATATGCTGTGAACCCTC 635  
| | | | | | | | | | | | | | | | | | | | | |  
Db 585 CGTGCGGGTGCCTGATCAAACAACATGATGCCGCAAAATCAAAGTGATCGCCGTAGAAGCGGA 644  
| | | | | | | | | | | | | | | | | | | | | |  
QY 636 AAATGCAGATGACTGCTACCAAGTCGAAGCTGAAGGGGAACATGATGCCCAATCTTTATCC 695  
| | | | | | | | | | | | | | | | | | | | | |  
Db 645 AGACTCCGCTGCTGCTGAAGCAGCCCTGGATCGGGTTCATCCGGTTGATCTGCCCGCGGT 704  
| | | | | | | | | | | | | | | | | | | | | |  
QY 696 TCCAGAACCATAGCAGTAGGTGTCAAATTCAGACATTGGCTTGAACACACTGCGCTATTAT 755  
| | | | | | | | | | | | | | | | | | | | | |  
Db 705 AGGGCTATTTCGTGAAGCGTAGCGGTAAACGCATCGTGACGAAACCTCCCTCGTTATG 764  
| | | | | | | | | | | | | | | | | | | | | |  
QY 756 CAGGGACCTTGTGGATGATATCTTCACTCTCACAGAGGATGAAT 800  
| | | | | | | | | | | | | | | | | | | | | |  
Db 765 CCAGGAGTATCTCGACGACATCATCCCTCGATAGCGATCGCAT 809  
| | | | | | | | | | | | | | | | | | | | | |

RESULT 7  
US-08-628-039-8  
; Sequence 8, Application US/08628039  
; Patent No. 5942660  
; GENERAL INFORMATION:  
; APPLICANT: Gruys, Kenneth J.  
; APPLICANT: Mitsky, Timothy A.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stark, David M.  
; APPLICANT: Hinchee, Maud A. W.  
; APPLICANT: Clemente, Thomas E.  
; APPLICANT: Connor-Ward, Dannette V.  
; APPLICANT: Fedele, Mary J.  
; APPLICANT: Fry, Joyce E.  
; APPLICANT: Howe, Arlene R.  
; APPLICANT: Roman, Renee J.  
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and  
; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyval  
; TITLE OF INVENTION: in Bacteria and plants  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gary M. Bond, Monsanto Company, A3SB  
; STREET: 800 No. 5942660th Lindbergh Boulevard  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63167  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/628,039  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE: 13-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bond, Gary  
; REGISTRATION NUMBER: 29,283  
; REFERENCE/DOCKET NUMBER: 38-21(13585)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314)694-3412  
; TELEFAX: (314)695-5435  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1545 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-628-039-8



US-08-673-388-5  
; Sequence 5, Application US/08673388  
; Patent No. 5958745

## GENERAL INFORMATION:

APPLICANT: Gruys, Kenneth J.  
APPLICANT: Mitsky, Timothy A.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Slater, Steven C.  
APPLICANT: Padgett, Stephen R.  
APPLICANT: Stark, David M.  
APPLICANT: Hinchee, Maud A. W.  
APPLICANT: Clemente, Thomas E.  
APPLICANT: Connor-Ward, Dannette V.  
APPLICANT: Fedele, Mary J.  
APPLICANT: Fry, Joyce E.  
APPLICANT: Howe, Arlene R.  
APPLICANT: Rozman, Renee J.

TITLE OF INVENTION: Methods of Optimizing Substrate Pools and

TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalerate

TITLE OF INVENTION: in Bacteria and Plants

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Charles E. Cohen, Monsanto Company, BBAF

STREET: 700 Chesterfield Parkway No. 5958745th

CITY: St. Louis

STATE: Missouri

COUNTRY: USA

ZIP: 63198

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/673,388

FILING DATE: 13-MAR-1996

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Cohen, Charles E.

REGISTRATION NUMBER: 34,565

REFERENCE/DOCKET NUMBER: 38-21(13618)A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (314)537-6224

TELEFAX: (314)537-6047

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1545 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-673-388-5

Query Match 5.68; Score 57; DB 2; Length 1545;

Best Local Similarity 43.68; Pred. No. 1.9e-08;

Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Qy 216 AGAAGGAGCGCAAGCTGTGTACTACAGCAGTGGAAACCATCGCCAGCGCTCTCAC 275

Db 225 AGACAGAAAGCGCACGGCTGATCACTGCTTTCGGGGTAACACCGCGGCGTCCG 284

Qy 276 CTATGCTGCCAAATGGAAGAAATTCCTGCTTATATTGTGTGTCGCCCGACAGCTCCAGA 335

Db 285 GTTTCTCTCGCGGGTTAGCGGTGAAGCCCTGATCGTTATGCCACCGCCACCGCGCA 344

Qy 336 CTGTAAATAACTTCAATACAGCCTACGAGCGTCAATTGTATATGTGAACCTAGTGA 395

Db 345 CATCAAAAGTACCGCGTTCGCGGGCTTCGGCGGCAAGTCTGCTCCACCGCGGCAACTT 404

Qy 396 TGAGTCCAGAGAAAATGTTCAAAAACAGTTACAGAAGAAACAGAGGCGATCATGGTACA 455

Db 405 TGATGAAGCGCAACGCAAGGATCGAAGCTGTCTACACGACGAGGGGTTACCTGGGTGCC 464

Qy 456 TCCCAACGAGGAGCCTGCAGTGATAGTGTGACAGAGGACAATTCCTGTGAAGTGTGAA 515  
Db 465 GCCCTTCGACCATCCGATGGTGATTGCGGGCAAGGACGCGTGGCTGGAACCTCCA 524  
Qy 516 CCAGGTTCCCTTGGTGGATGCACCTGGTGGTACCTGTAGTGTAGGAGGAGGAATGCTTGG 575  
Db 525 GCAGGACGCCCATCTCGACCGCGTATTGTGCCAGTGGCGGGCGGCTGTGGCTGCTTG 584  
Qy 576 AATAGCAATTACAGTTAAGGCTCTGAAACCTTAGTGTGAAGGTATATGCTGCTGAACCTC 635  
Db 585 CGTGGCGGTCTGATCAACAACATGATGCCCAATCAAGTATCCCGCTAGAACGCGA 644  
Qy 636 AATGTCAGATGCTGTACCAAGTCCAAAGCTGAAGGGGAAACTGATGCCCAATCTTTATPC 695  
Db 645 AGACTCGGCTGCTGCTGAAAGCAGCGCTGGATCGGGTTCATCCGCTTGATCTGCCCGCGT 704  
Qy 696 TCCAGAAACCATACGACATGGTGTCAAAATCCAGCATTTGGCTTGAACACCTGGCCCTATTAT 755  
Db 705 AGGCTATTGCTGAAGCGGTAGCGGTAAACGATCGGTGACGAAACCTTCCTGTTATG 764  
Qy 756 CAGGACCTTGTGGATGATATCTTCACTGTACACAGAGGATGAAAT 800  
Db 765 CCAGGAGTATCTCGACGACATCATCACCGTCCGATAGCGATCGCAT 809

## RESULT 10

US-08-673-388-7

; Sequence 7, Application US/08673388

; Patent No. 5958745

; GENERAL INFORMATION:

APPLICANT: Gruys, Kenneth J.

APPLICANT: Mitsky, Timothy A.

APPLICANT: Kishore, Ganesh M.

APPLICANT: Slater, Steven C.

APPLICANT: Padgett, Stephen R.

APPLICANT: Stark, David M.

APPLICANT: Hinchee, Maud A. W.

APPLICANT: Clemente, Thomas E.

APPLICANT: Connor-Ward, Dannette V.

APPLICANT: Fedele, Mary J.

APPLICANT: Fry, Joyce E.

APPLICANT: Howe, Arlene R.

APPLICANT: Rozman, Renee J.

TITLE OF INVENTION: Methods of Optimizing Substrate Pools and

TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyval

TITLE OF INVENTION: in Bacteria and Plants

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Charles E. Cohen, Monsanto Company, BBAF

STREET: 700 Chesterfield Parkway No. 5958745th

CITY: St. Louis

STATE: Missouri

COUNTRY: USA

ZIP: 63198

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/673,388

FILING DATE: 13-MAR-1996

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Cohen, Charles E.

REGISTRATION NUMBER: 34,565

REFERENCE/DOCKET NUMBER: 38-21(13618)A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (314)537-6224

TELEFAX: (314)537-6047

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1545 base pairs



Db 765 CCAGGAGTATCTCGACGACATCATCACCGTCGATAGCGAT 809

## RESULT 12

US-08-614-877-1

; Sequence 1, Application US/08614877

; Patent No. 5959179

; GENERAL INFORMATION:

; APPLICANT: Gruys, Kenneth J.

; APPLICANT: Milsky, Timothy A.

; APPLICANT: Kishore, Ganesh M.

; APPLICANT: Slater, Steven C.

; APPLICANT: Padgett, Stephen R.

; APPLICANT: Stark, David M.

; APPLICANT: Hinchee, Maud A. W.

; APPLICANT: Clemente, Thomas E.

; APPLICANT: Connor-Ward, Dannette V.

; APPLICANT: Fedele, Mary J.

; APPLICANT: Fry, Joyce E.

; APPLICANT: Howe, Arlene R.

; APPLICANT: Rozman, Renee J.

; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and

; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalerate

; TITLE OF INVENTION: in Bacteria and Plants

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Charles E. Cohen, Monsanto Company, BB4F

; STREET: 700 Chesterfield Parkway No. 5959179th

; CITY: St. Louis

; STATE: Missouri

; COUNTRY: USA

; ZIP: 63198

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/614,877

; FILING DATE: 13-MAR-1996

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Cohen, Charles E.

; REGISTRATION NUMBER: 34,565

; REFERENCE/DOCKET NUMBER: 38-21(10695)A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (314)537-6224

; TELEFAX: (314)537-6047

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1545 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-614-877-1

Query Match 5.6% Score 57; DB 2: Length 1545;

Best Local Similarity 43.6% Pred. No. 1.9e-08;

Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Qy 216 AGAAGGAGCCGAAAGCTGTTACTACAGCAGTGGAACCATGGCCAGGCTCTCAC 275

Db 225 AGAACAGAAAGCGACGGCGTGATCATCTGCTCGGGTAACACCGCGAGGCGTCGC 284

Qy 276 CTATGTCGCAATTTGAAGAAATTCCTGCTTATATTTGGTGGTCCCCAGACAGCTCAGA 335

Db 285 GTTTTCTTCGCGGGTTAGGCGTGAAGGCCCTGATGCTTATGCCAACCCACCGCGGA 344

Qy 336 CTGTAAAAAATTCGATACAGCCCTACGAGCGCTCAATTTGTATATCTGTGAACCTAGTGA 395

Db 345 CATCAAAAGTCGACCGGTGCGCGGCTTCGCGCGGAGAGTGCTGCTCCACGGCGCACTT 404

Qy 396 TGAGTCCAGAGAAAATGTTGCAAAAAAGATTACAGAAAGAAACAGAGGCATCATGGTACA 455  
Db 405 TGATGAAGCGAAACGCAACCGATCGAACTGTACACAGCAGCAGGGGTTACCTGGGTGCC 464  
Qy 456 TCCCAACAGGAGCGCTGCTAGCTGATGCTGACAAAGGACAAATGCCCCCTGGAAGTGTCTGAA 515  
Db 465 GCGGTTTCGACCATCCGATGCTGATGTCGCGGCAAGGCACGCTGCGGCTGGAACCTGCTCCA 524  
Qy 516 CCAGGTTCCCTTTGGTGGATGCACCTGGTGGTACCTGTAGGTGGAGGAGGAATGCTTGCTGG 575  
Db 525 GCAGGAGCGCCATCTCGACCGCGTATTTGTGCCAGTCGGCGGCGGCTGCGTGGCTTGTG 584  
Qy 576 AATAGCAATTTACAGTTTAAGGCTCTGAAACCTTAGTGTGAAGGTATATGCTGCTGAACCCCTC 635  
Db 585 CGTGGCGGTGCTGATCAAAACAACTGATGCCGCCAAATCAAAAGTGATGCCCGTAGAAGCGGA 644  
Qy 636 AATGTCAGATGACTGCTACCAAGCTCAAGCTGAAGGGAAGAACTGATGCCCAATCTTTATCC 695  
Db 645 AGACTCCGCTGCTGCTGAAAGCAGCGCTGATGCGGGTCACTCCGTTGATCTGCGCGCGCT 704  
Qy 696 TCCAGAAACCATAGCAGATGCTGTCAAATCCAGCATTTGGCTTGAACACCTGGCCTATTAT 755  
Db 705 AGGCTATTTGCTGAAGCGGTAGCGGTAAACCATCATCCGTCGATAGGATGCGAT 764  
Qy 756 CAGGAGCCTTTGGTGGATGATATCTTACTGTACAGAGGATGAAAT 800  
Db 765 CCAGGAGTATCTCGACGACATCATCACCGTCGATAGCGAT 809

## RESULT 13

US-08-614-877-5

; Sequence 5, Application US/08614877

; Patent No. 5959179

; GENERAL INFORMATION:

; APPLICANT: Gruys, Kenneth J.

; APPLICANT: Milsky, Timothy A.

; APPLICANT: Kishore, Ganesh M.

; APPLICANT: Slater, Steven C.

; APPLICANT: Padgett, Stephen R.

; APPLICANT: Stark, David M.

; APPLICANT: Hinchee, Maud A. W.

; APPLICANT: Clemente, Thomas E.

; APPLICANT: Connor-Ward, Dannette V.

; APPLICANT: Fedele, Mary J.

; APPLICANT: Fry, Joyce E.

; APPLICANT: Howe, Arlene R.

; APPLICANT: Rozman, Renee J.

; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and

; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyval

; TITLE OF INVENTION: in Bacteria and Plants

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Charles E. Cohen, Monsanto Company, BB4F

; STREET: 700 Chesterfield Parkway No. 5959179th

; CITY: St. Louis

; STATE: Missouri

; COUNTRY: USA

; ZIP: 63198

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/614,877

; FILING DATE: 13-MAR-1996

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Cohen, Charles E.

; REGISTRATION NUMBER: 34,565

; REFERENCE/DOCKET NUMBER: 38-21(10695)A

; TELECOMMUNICATION INFORMATION:

TELEPHONE: (314)537-6224  
TELEFAX: (314)537-6047  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1545 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-614-877-5

Query Match	5.6%	Score 57:	DB 2:	Length 1545;
Best Local Similarity	43.6%;	Pred. No. 1.9e-08;		
Matches 255;	Conservative 0;	Mismatches 330;	Indels 0;	Gaps 0;
QY	216	AGAAAGGAAGCCGAAAGCTCTGTTACTACACGACAGTGGAAACCATGGCCAGGCTCTCAC	275	
Db	225	AGAACGAAGCGACCGCGTGATCACTGCTTCGCGGTAAACACCGCAGGCGGTGCG	284	
QY	276	CTATGCTGCCAAATTTGAAGGAATTCCTGCTTATATTGTGTGCCGCCACAGACTCCAGA	335	
Db	285	GTWTTCTCTCGCGGTTAGSCGTGAAGGCCCTGATCGTTATGCCACCCGCCACGCCGA	344	
QY	336	CTGTAAAAACTTGCRAATACAAAGCCTACGAGCGGTCAATTGTATCTGTGAACCTAGTGA	395	
Db	345	CATCAAAGTCGACCGCGTGGCGGCTTCGCGCGCGAAGTGTCTCCACGGCGGAACTT	404	
QY	396	TGAGTCCAGAGAAAATGTTTCCAAAAGAGATTACAGAAGAAACAGAGGCATCATGGTACA	455	
Db	405	TGATGAAGCGAAACGCNAAGCGATCGNACTGTACACGACAGGGGTTACCTGGGTGCC	464	
QY	456	TCCCAACAGGAGCGCTGCAGTAGCTAGCTGACAAAGGACAAATTCCTCGGAAGTCTGAA	515	
Db	465	GCGGTTTCGACATCCGATGTTGATTGCGGGCAAGGCACGTGCGCTGTGAACCTGCCA	524	
QY	516	CCAGGTTCTCTTGTGTCATGTCAGTGTGTGCTAGCTGTAGGTGGAGGAGGAATGCTTGTGG	575	
Db	525	GCAGGAGGCCATCTCGACCGCGTATTGTGCCAGTCGCGCGCGGCTGTGGCTGTG	584	
QY	576	AATAGCAATTACAGTTAAGCTCTGAAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTC	635	
Db	585	CGTGGCGGTGCTGATCAAAACAATGATGCCGCAAAATCAAAGTGATGCCCGTAGAAGCGGA	644	
QY	636	AAATGCAGATGACTGCTACCAAGTCCAAAGCTGAAGGGGAAACTGATGCCCAATCTTTATCC	695	
Db	645	AGACTCCGCGTCCCTGAAAGCAGCGCTGGATCGGGTCACTCCGCTGTGATCTGCCGCGGT	704	
QY	696	TCCAGAAACCATAGCAGATGGTTCAAATCCAGCATTTGGCTTGAACACCTTGGCTATTAT	755	
Db	705	AGGGCTATTGTGTGAAGGCGTACGCGGTAAACGCATCGGTGACGAACCTTCGGTTATG	764	
QY	756	CAGGGACCTTGTGGATGATATCTTCACTGTCACAGAGGATGAAAT	800	
Db	765	CCAGGAGTATCTCGACGACATCATCAACCGTCGATAGCGATGCCAT	809	

RESULT 14  
US-08-614-877-7  
Sequence 7, Application US/08614877  
Patent No. 5959179  
GENERAL INFORMATION:  
APPLICANT: Gruys, Kenneth J.  
APPLICANT: Mitsky, Timothy A.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Slater, Steven C.  
APPLICANT: Padgett, Stephen R.  
APPLICANT: Stark, David M.  
APPLICANT: Hincee, Maud A. W.  
APPLICANT: Clemente, Thomas E.  
APPLICANT: Connor-Ward, Dannel  
APPLICANT: Fedele, Mary J.  
APPLICANT: Fry, Joyce E.  
APPLICANT: Howe, Arlene R.

QY 696 TCCAGAACCATACGACATGGTCAAAATCCAGCATTTGGCTTGAACACCTGGCCATTAT 755  
Db 705 AGGCTATTGGCTGAAGCGGTAGCGGTAAACGCGATCGGTGACGAAACCTTCGGTTTATG 764  
QY 756 CAGGACCTTGTGATGATATCTTCACTGTGCACAGAGGATGAAAT 800  
Db 765 CCAGGACTATCTGACGACATCATCACCGTCGATAGCGATGCGAT 809

## RESULT 15

US-08-614-877-8  
; Sequence 8, Application US/08614877  
; Patent No. 5959179  
; GENERAL INFORMATION:  
; APPLICANT: Gruys, Kenneth J.  
; APPLICANT: Mitsky, Timothy A.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stark, David M.  
; APPLICANT: Hinchey, Maud A. W.  
; APPLICANT: Clemente, Thomas E.  
; APPLICANT: Connor-Ward, Dannette V.  
; APPLICANT: Fedele, Mary J.  
; APPLICANT: Fry, Joyce E.  
; APPLICANT: Howe, Arlene R.  
; APPLICANT: Rozman, Renee J.  
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and  
; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalera  
; TITLE OF INVENTION: in Bacteria and Plants  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Charles E. Cohen, Monsanto Company, B44F  
; STREET: 700 Chesterfield Parkway No. 5959179th  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/614,877  
; FILING DATE: 13-MAR-1996  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cohen, Charles E.  
; REGISTRATION NUMBER: 34,565  
; REFERENCE/DOCKET NUMBER: 38-21(10695)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314)537-6224  
; TELEFAX: (314)537-6047  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1545 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-614-877-8

Query Match 5.6%; Score 57; DB 2; Length 1545;  
Best Local Similarity 43.6%; Pred. No. 1.9e-08;  
Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;  
QY 216 AGAAGAACCCGAAAGCTTTGTTACTACAGCAGTGGAAACCATGGCCAGGCTCTCAC 275  
Db 225 AGAACAGAACGCGACGCGGTGATCACTGCTTGGCGGTAAACACGCGCAGGCGGTGCG 284  
QY 276 CTATGCTGCCAAATTGAAGGAATTCCTGCTTATATGTCGCCCGCAGACAGCTCCAG 335

Db 285 GTTTTCTTCTGCGCGGTTAGCGGTGAAGGCCCTGATCGTTATGCCAACCCGCCACCGCGCA 344  
QY 336 CTGTAAAAAATTGCAATACAAAGCCTACGGAGCGCTCAATTTGTATACCTGTGAACCTAGTGA 395  
Db 345 CATCAAAGTCGACCGGCTTCGCGCGGGAAGTGTCTCCACGCGCGAACTT 404  
QY 396 TGAGTCCAGAGAAAATTTGCAAAAAGAGTTTCAAGAAACAGAACAGAGGCGATCATGTGTACA 455  
Db 405 TGATGAAGCGAAACGCAAGCGATCGAACTGTTCACAGCAGCAGGGGTTACCTGGGTGCC 464  
QY 456 TCCCAACACGAGGCGCTCGACGTAGCTGGACAAGGGACAAATTCGCCCTGGAAGTGTGAA 515  
Db 465 GCCGTTCCGACCATCCGATGGTGTGTCGGGCAAGGACCGCTGGCGCTGGAACCTGCTCCA 524  
QY 516 CCAGGTTCCCTTTGGTGGATGCACCTGGTGGTACCTGTAGTGGAGGAGGAATGTGCTGG 575  
Db 525 GCAGGACGCGCATCTCGACCGCGTATTTGTGCCAGTCGCGCGGCGGTCTGGCTGCTTG 584  
QY 576 AATAGCAATTTACAGTTAAGGCTCTGAAAACCTAGTGTGAAGGTATATGCTGCTGAACCTC 635  
Db 585 CGTGGCGGTGCTGATCAAAACAACTGATGCCGCAAAATCAAAGTGTATCCCGGTAGAAGCGGA 644  
QY 636 AATGTCAGATGACTGTCTACCAGTCCAAGCTGAAGGGGAAACTGATGCCCAATCTTATCC 695  
Db 645 AGACTCCGCTGCTGAAAGCAGCGCTGGATGGGGGTATCCCGGTTGATCTGCCCGCGGT 704  
QY 696 TCCAGAAACCATAGCAGATGGTGTCAAAATCCAGCATTTGGCTTGAACACCTGGCGCTATTAT 755  
Db 705 AGGCTATTTGCTCAAGCGCTAGCGGTAAAACGCGATCGGTGACCAAACTTCCGTTTATG 764  
QY 756 CAGGACCTTGTGGATGATATCTTCACTGTGCACAGAGGATGAAAT 800  
Db 765 CCAGGAGTATCTCGACGACATCATCACCGTCGATAGCGATGCGAT 809

Search completed: June 24, 2003, 06:24:42

Job time : 81 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 03:27:09 ; Search time 2668 seconds  
(without alignments)  
11104.444 Million cell updates/sec

Title: US-09-889-609B-1  
Perfect score: 1018  
Sequence: 1 atgtgtgtcagctactgcatt.....ttaccagacggtctgttttaa 1018

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rtd.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1011.2	99.3	1197	10	AF148321	AF148321 Mus muscu
2	1011.2	99.3	1411	10	BC011164	BC011164 Mus muscu
3	807.8	79.4	1020	6	AX224420	AX224420 Sequence
4	807.8	79.4	1023	9	AF034081	AF034081 Homo sapi
5	807.8	79.4	1166	9	AF169974	AF169974 Homo sapi
6	807.8	79.4	1770	6	AX224418	AX224418 Sequence
7	807.8	79.4	2477	9	AK023169	AK023169 Homo sapi
8	806.2	79.2	1336	6	AX259575	AX259575 Sequence
9	664.6	65.3	2264	9	HS080540	AL834378 Homo sapi
10	257.2	25.3	184151	10	AL604066	AL604066 Mouse DNA
11	257.2	25.3	205910	2	AC044847	AC044847 Mus muscu
12	250.8	24.6	3379	10	BC031531	BC031531 Mus muscu
13	212.2	20.8	101814	2	AC122087	AC122087 Rattus no
14	181.8	17.9	3807	9	BC019090	BC019090 Homo sapi
15	181.8	17.9	4140	9	AK026565	AK026565 Homo sapi
16	181.8	17.9	123454	9	HSBC17A99	AL450226 Homo sapi
17	181.8	17.9	221909	2	AC021705	AC021705 Homo sapi
18	151.4	14.9	40005	8	SPCC320	AL022245 S.pombe c
19	151.4	14.9	42689	8	SPCC330	AL031603 S.pombe c
20	142	13.9	295350	1	AP004826	AP004826 Staphyloc
21	140.2	13.8	4107	9	AB037822	AB037822 Homo sapi
22	130	12.8	6525	8	SCYKL218C	228218 S.cerevisia
23	130	12.8	24577	8	SC60RF	X75951 S.cerevisia
24	124.4	12.2	303750	1	AP003133	AP003133 Staphyloc
25	124.4	12.2	346900	1	AP003362	AP003362 Staphyloc
26	119.4	11.7	166050	1	AL646085	AL646085 Ralstonia
27	117.8	11.6	157609	2	AC095280	AC095280 Rattus no
28	116.8	11.5	16931	1	AE001716	AE001716 Thermotoga
29	114.4	11.2	298050	1	AP003189	AP003189 Clostridi
30	110.4	10.8	14449	1	AE004696	AE004696 Pseudomon
31	110.2	10.8	291050	1	AP000982	AP000982 Sulfolobu
32	108.6	10.7	35437	1	SCF43A	AL096837 Streptomy
33	107.4	10.6	2186	1	ECTDC	M21312 E.coli tdc
34	107.4	10.6	6295	1	ECTDCRAB	X14430 E.coli tdc
35	107.4	10.6	10516	1	AE000393	AE000393 Escherich
36	107.4	10.6	13979	1	AE005540	AE005540 Escherich
37	107.4	10.6	110000	1	ECOUW67.0	U18997 Escherichia
38	107.4	10.6	307962	1	AP002564	AP002564 Escherich
39	101	9.9	20878	1	AE008849	AE008849 Salmonell
40	97.8	9.6	258050	1	AL627278	AL627278 Salmonell
41	97.4	9.6	137	11	G32814	G32814 A009S27 Hum
42	97	9.5	303040	1	AP000991	AP000991 Thermopla
43	95.4	9.4	268200	1	AP000061	AP000061 Aeropyrum
44	94.8	9.3	10163	1	AE006661	AE006661 Sulfolobu
45	93.4	9.2	9990	1	AE010646	AE010646 Fusobacte

ALIGNMENTS

RESULT 1  
AF148321  
LOCUS AF148321 1197 bp mRNA linear ROD 18-NOV-1999  
DEFINITION Mus musculus serine racemase mRNA, complete cds.  
ACCESSION AF148321  
VERSION AF148321.1 GI:6448864  
KEYWORDS Mus musculus.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 1197)  
AUTHORS Wolosker,H., Blackshaw,S. and Snyder,S.H.  
TITLE Serine racemase: a glial enzyme synthesizing D-serine to regulate glutamate-N-methyl-D-aspartate neurotransmission

Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13409-13414 (1999)  
20027561  
MEDLINE  
PUBMED  
10557334  
REFERENCE  
2 (bases 1 to 1197)  
Wolosker, H.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (03-MAY-1999) Neuroscience, Johns Hopkins University, 725  
JOURNAL North Wolfe St., WBSB 806, Baltimore, MD 21205, USA  
Location/Qualifiers  
FEATURES  
source  
1. 1197  
/organism="Mus musculus"  
/strain="BALB/c"  
/db\_xref="taxon:10090"  
66. 1085  
/EC\_number="5.1.1.17"  
/functions="catalyzes L-serine to D-serine racemization"  
/codon\_start=1  
/product="serine racemase"  
/protein\_id="AAF08701.1"  
/db\_xref="GI:6448865"  
/translaton="MCAQYCTGSPADVEKHAHINIQDSIHLTPVLTSILNOIAGRNLF  
KCELFQKTSFKIRLNAHNRGLIPDTPPEKPKAVVTHSSNHCQALTYAAKLEGI  
YIVVPQAPNCKKLAIOAGASIVYCDPSDESREKVTQRIOMETEGLIVHPNPEPAVI  
AGQSTALEVNLQPLVDVALVPVGGGGMVAGIATIKALKPSKVTAAPSPNADDCY  
QSKLKGELTNLNPEDTADGVKSSIGLNTVIRLDVDDVFTVEDIEIKATQLVWVG  
RMKLLIEPTAGVALAAVLVSOFHQTVPSEVKNVCIVLSGGNVDLTSLNWWGQAPRPAY  
QTVSV"  
BASE COUNT 345 a 283 c 283 g 286 t  
ORIGIN  
Query Match 99.3%; Score 1011.2; DB 10; Length 1197;  
Best Local Similarity 99.7%; Pred. No. 1.9e-267;  
Matches 1013; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 ATGTGTGCTCAGTACTGCATCTCCTTGTGCTGATGTTGAAAAGCTCATATCAACATTCAA 60  
Db 66 ATGTGTGCTCAGTACTGCATCTCCTTGTGCTGATGTTGAAAAGCTCATATCAACATTCAA 125  
Qy 61 GACTCTATCACCTCACCCAGTCTTAACAGCTCCATTTTGAATCAAAATAGCAGGCGC 120  
Db 126 GACTCTATCACCTCACCCAGTCTTAACAGCTCCATTTTGAATCAAAATAGCAGGCGC 185  
Qy 121 AATCTTTTCTTCAAATGTGAGCTCTTCCAGAAAACCTGGTCTTTTAAAGATTCGAGGTGCC 180  
Db 186 AATCTTTTCTTCAAATGTGAGCTCTTCCAGAAAACCTGGTCTTTTAAAGATTCGAGGTGCC 245  
Qy 181 CTTAATGCCATCAGAGCGTTAATTCCTGACAGCCGCAAGAGAGCCCAAGCCGTAGTT 240  
Db 246 CTTAATGCCATCAGAGCGTTAATTCCTGACAGCCGCAAGAGAGCCCAAGCCGTAGTT 305  
Qy 241 ACTCACAGCAGCGGAAACACATGGCCAGCTCTCAACCTATGCTGCTAAACTTGAAGAGGAAT 300  
Db 306 ACTCACAGCAGCGGAAACACATGGCCAGCTCTCAACCTATGCTGCTTAACTTGAAGAGGAAT 365  
Qy 301 CCTGCTTACATTTGTGGTTCCCAACAGCTCCCAACCTGCAAGAACTGGCAATCCAAAGCC 360  
Db 366 CCTGCTTACATTTGTGGTTCCCAACAGCTCCCAACCTGCAAGAACTGGCAATCCAAAGCC 425  
Qy 361 TATGGAGATCGATAGTATATCTGTGACCCAAAGTGACGAGTCCAGAGAAAAGGTCACTCAA 420  
Db 426 TATGGAGATCGATAGTATATCTGTGACCCAAAGTGACGAGTCCAGAGAAAAGGTCACTCAA 485  
Qy 421 AGAATTATGAAGAAACAGAGGATCTTGGTCCATCCCAACAGGAGCCTGCAGTGATA 480  
Db 486 AGAATTATGAAGAAACAGAGGATCTTGGTCCATCCCAACAGGAGCCTGCAGTGATA 545  
Qy 481 GCTGGCAAGGAACAATTTGCCCTTGGAAAGTGTGAAACAGGTTCCCTTGGTGTAGATGCACCTG 540  
Db 546 GCTGGCAAGGAACAATTTGCCCTTGGAAAGTGTGAAACAGGTTCCCTTGGTGTAGATGCACCTG 605  
Qy 541 GTGTTACAGTAGGAGGAGGAGGAATGGTTGCTGGAATAGCCATTACAATTAAGGCCCTG 600  
Db 606 GTGTTACAGTAGGAGGAGGAGGAATGGTTGCTGGAATAGCCATTACAATTAAGGCCCTG 665

QY	601	AAACCTAGTGTGAAGGTATACGCTGTGTGAGCCCTCGAATGCAGATGACTGCTACCACTCT	660					
Db	666	AAACCTAGTGTGAAGGTATACGCTGTGTGAGCCCTCGAATGCAGATGACTGCTACCACTCT	725					
QY	661	AAACTGAAAGGAAACCTGACCCCAATCTTCCTCCAGAAACCATAGCAGATGGTGTC	720					
Db	726	AAACTGAAAGGAAACCTGACCCCAATCTTCCTCCAGAAACCATAGCAGATGGTGTC	785					
QY	721	AAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATTAAGAGACCTTTGTGGATGATGCTTC	780					
Db	786	AAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATTAAGAGACCTTTGTGGATGATGCTTC	845					
QY	781	ACTGTCCAGCAAGTGAATCAAGTATGCAACCCAGCTGTGTGGGGGGAAGATGAACCTG	840					
Db	846	ACTGTCCAGCAAGTGAATCAAGTATGCAACCCAGCTGTGTGGGGGGAAGATGAACCTG	905					
QY	841	CTCATTGAGCCGACTGCTGGCGTGGCACTGGCTGCAGTGTCTCAGCAGATTTCCAAACA	900					
Db	906	CTCATTGAGCCGACTGCTGGCGTGGCACTGGCTGCAGTGTCTCAGCAGATTTCCAAACA	965					
QY	901	GTCTCTCCAGAAGTAAAGAACGCTCTGCATTGTACTCAGTGGGGGGAATGTAGACCTTAACC	960					
Db	966	GTCTCTCCAGAAGTAAAGAACGCTCTGCATTGTACTCAGTGGGGGGAATGTAGACCTTAACC	1025					
QY	961	TCCTGACTGGTGGGGGAGCTGTAACGCCAGCTCTTACCAGACGGTCTGTTT	1016					
Db	1026	TCCTGACTGGTGGGGGAGCTGTAACGCCAGCTCTTACCAGACGGTCTGTTT	1081					
RESULT 2								
LOCUS	BC011164	1411 bp mRNA linear	ROD 07-AUG-2002					
DEFINITION	Mus musculus, Similar to serine racemase, clone MGC:118670							
IMAGE	IMAGE:4195695, mRNA, complete cds.							
ACCESSION	BC011164							
VERSION	BC011164.1	GI:15029880						
KEYWORDS	MGC.							
SOURCE	house mouse.							
ORGANISM	Mus musculus							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							
AUTHORS	1 (bases 1 to 1411)							
TITLE	Strausberg, R.							
JOURNAL	Direct Submission							
	Submitted (25-JUL-2001) National Institutes of Health, Mammalian							
	Gene Collection (MGC), Cancer Genomics Office, National Cancer							
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,							
	USA							
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>							
COMMENT	Contact: MGC help desk							
	Email: <a href="mailto:cgabbs@mail.nih.gov">cgabbs@mail.nih.gov</a>							
	Tissue Procurement: Jeffrey E. Green, M.D.							
	cDNA Library Preparation: Life Technologies, Inc.							
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)							
	DNA Sequencing by: Baylor College of Medicine Human Genome							
	Sequencing Center							
	Center code: BCM-HGSC							
	Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a>							
	Contact: <a href="mailto:amg@bcm.tmc.edu">amg@bcm.tmc.edu</a>							
	Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,							
	Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,							
	Richards, S., Gibbs, R.A.							
Clone distribution: MGC clone distribution information can be found								
through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>								
Series: IRAK Plate: 24 Row: m Column: 6								
This clone was selected for full length sequencing because it								
passed the following selection criteria: matched mRNA gi: 7305520.								
FEATURES	Location/Qualifiers							
source	1..1411							
	/organism="Mus musculus"							
	/db_xref="taxon:10090"							



```
QY 361 TATGGAGCATCGATAGTATATCTGTGACCCCAAGTGACGAGTCCAGAGAAAGGTCACCTCAA 420
Db 361 TACGGAGCGTCATTTGTATCTGTGAACCTAGTGTAGTCCAGAGAAATGTTGCAAAA 420
QY 421 AGAATTATGCAAGAACAAGAGCATCTTGGTCCATCCCAACCAAGGAGCGCTGCAGTGATA 480
Db 421 AGATTACAGAAGAACAAGAGCATCATGTTACATCCCAACCAAGGAGCGCTGCAGTGATA 480
QY 481 GCTGGCAAGGAACAATTCCTGGAGTGCTCAACCAAGCTTCCCTGGTGTAGATGCACCTG 540
Db 481 GCTGGCAAGGAACAATTCCTGGAGTGCTCAACCAAGCTTCCCTGGTGTAGATGCACCTG 540
QY 541 GTGGTACCAAGTAGGAGGAGGAATGTTGCTGGAAATAGCCATTAACAATTAAGGCCCTG 600
Db 541 GTGGTACCTCTAGGTGGAGGAGGAATGTTGCTGGAAATAGCAATTAAGGCTCTG 600
QY 601 AAACCTAGTGTGAAGTATACGCTGTGAGCCCTCGAATGCAGATGACTGCTACCAAGTCT 660
Db 601 AAACCTAGTGTGAAGTATATGCTGTGAACCTCAATGCAGATGACTGCTACCAAGTCT 660
QY 661 AAACCTGAAGGAGCAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGCTGTC 720
Db 661 AAACCTGAAGGAGCAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGCTGTC 720
QY 721 AAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATAGAGACCTTGTGGATGATGCTTC 780
Db 721 AAATCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGACCTTGTGGATGATGCTTC 780
QY 781 ACTGTCAACCAAGATGAATCAAGTATGCAACCCAGCTGTGTGGGGGAGAAATGAACATG 840
Db 781 ACTGTCAACCAAGATGAATTAAGTGTGAACCCAGCTGTGTGGGGGAGAAATGAACATG 840
QY 841 CTCATTGAGCGGACTGTGGCTGGCACTGGCTGTCAGTGTCTCAGCATTTCCAAACA 900
Db 841 CTCATTGAACTACACTGGTGTGGAGTGGCTGTCTGCTGTCTCAGCATTTCCAAACT 900
QY 901 GTCTCTCCAGAAGTAAGAACGCTCTCATTTACTAGTGGGGGAAATGTAGACCTAA-- 958
Db 901 GTTTCCTCCAGAAGTAAGAACATTTGTTGCTCAGTGGTGAATGTAGACTTTAACC 960
QY 959 -CCTCCCTGACCTGGCTGGGAGGCTGAACGCCAGCTCTTACCAGAGCGTCTGTT 1016
Db 961 TCCTCCATACTTGGGTGAAGCAGGCTGAAAGCCAGCTCTTATCAGTGTGTTCTGT 1019

RESULT 4
LOCUS AY034081 1023 bp mRNA linear PRI 01-JUN-2002
DEFINITION Homo sapiens serine racemase mRNA, complete cds.
ACCESSION AY034081
VERSION AY034081.1 GI:21307620
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1023)
AUTHORS Xia.M., Liu.Y. and Connolly.T.M.
TITLE Molecular cloning and expression of human serine racemase from human NT2N cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1023)
AUTHORS Xia.M., Liu.Y. and Connolly.T.M.
JOURNAL Direct Submission
TITL Submitted (08-MAY-2001) Pharmacology, Merck & Co., Sumneytown Pike, West Point, PA 19486, USA
FEATURES
source Location/Qualifiers
1..1023
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p13"
CDS
1..1023
```

```
/codon_start=1
/product="serine racemase"
/protein_id="AAK58495.1"
/db_xref="GI:21307621"
/translation="MCAQYCSIFADYVEKAHINIRDSIHLPVLTSSILNLTGRLNLF
KGELEFQTSFKIRGALNAVRSIVDALKERKPAVYTHSSGNHGQALTYAAKLEGP
YIVVPTAPDCKKLAIOAGSIVCEPDESRENVAKRTVEGIMVHPNCPDAVI
AGQGTALVNLQVLDALVVPVGGGMLAGIATVKAIPSKVYAAEPNSADDCY
QSKLKGKLPNLYPPETIADGVKSSIGLNTWPIRLVDIDFVTEDEIKCATQLWE
QKLLLEPTAGVGAVALSOHFQTVSPVKNICIVLGGNVDLTSSITWVQKAEPA
YOSVSV"
BASE COUNT 290 a 220 c 243 g 270 t
ORIGIN
Query Match 79.4%; Score 807.8; DB 9; Length 1023;
Best Local Similarity 87.7%; Pred. No. 2.1e-211;
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;
QY 1 ATGTGTGCTCAGTACTGTCATCTCTTTGCTGTGATGTTGAAAAAGCTCATATCAACATTC 60
Db 1 ATGTGTGCTCAGTATTGTCATCTCTTTGCTGTGATGTTGAAAAAGCTCATATCAACATTC 60
QY 61 GACTCTATCCACTCACCCAGTGTCAACAAGCTCCATTTTGAATCAAAATAGCAGGCGC 120
Db 61 GATTCTATCACCTCACACAGTGTCTAACAAGCTCCATTTTGAATCAAACTAACAGGCGC 120
QY 121 AATCTTTCTTCAAAATGTGAGCTCTTCCAGAAACTGGGTCTTTTAAAGATTCAGGTGCC 180
Db 121 AATCTTTCTTCAAAATGTGAACTCTTCCAGAAACAGGATCTTTTAAAGATTCGTGGTCT 180
QY 181 CTTAATGCCATCAGAGGCTTAATTTCTGTACACCCAGAGAGAGAGCCCAAGCCGCTAGTT 240
Db 181 CTCAATGCCCTCAGAGCTTGGTTCTGTGCTTTAGAAAGAGAGCCGAAAGCTGTGTT 240
QY 241 ACTCAGCAGCGGAAACCATAGCCAAAGCTCTCACCTATGCTGCTTAACTGGAAGAAAT 300
Db 241 ACTCAGCAGGTGGAACCATAGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGAAAT 300
QY 301 CTGTCTTACATTTGTGTTCCCAACAGCTGCCCACTGCAAGAACTGGCAATCCAGGCC 360
Db 301 CTGTCTTATTTGTGTTGCCAGACAGCTGCCAGACTGTAAANAATTTGCAATACAGGCC 360
QY 361 TATGAGCATCGATAGTATATCTGTACCCAAAGTGACGAGTCCAGAGAAAGGTCACTCAA 420
Db 361 TACGGAGCGTCAATTTCTATCTGTGAACCTAGTGTAGTCCAGACAGAAATGTTGCAAAA 420
QY 421 AGAATTATCGAAGAACAGAGGCTCTTGGTCCATCCCAACCAAGGAGCGCTGCAGTGATA 480
Db 421 AGAGTTTACAGAGAAACAGAGAGGCTCATGGTACATCCCAACCAAGGAGCGCTGCAGTGATA 480
QY 481 GCTGGCAAGGAACAATTCCTGGAGTGCTGAACCAAGCTTCCCTGGTGTAGATGCACCTG 540
Db 481 GCTGGCAAGGACAAATTCCTGGAGTGCTGAACCAAGCTTCCCTGGTGTAGATGCACCTG 540
QY 541 GTGGTACCAAGTAGGAGGAGGAATGTTGCTGGAAATAGCCATTAACAATTAAGGCCCTG 600
Db 541 GTGGTACCTCTAGGTGGAGGAGGAATGTTGCTGGAAATAGCAATTAAGGCTCTG 600
QY 601 AAACCTAGTGTGAAGTATACGCTGTGAGCCCTCGAATGCAGATGACTGCTACCAAGTCT 660
Db 601 AAACCTAGTGTGAAGTATATGCTGTGAACCTCAATGCAGATGACTGCTACCAAGTCT 660
QY 661 AAACCTGAAGGAGCAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGCTGTC 720
Db 661 AAACCTGAAGGAGCAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGCTGTC 720
QY 721 AAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATAGAGACCTTGTGGATGATGCTTC 780
Db 721 AAATCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGAGCTTGTGGATGATGCTTC 780
QY 781 ACTGTCAACCAAGATGAATCAAGTATGCAACCCAGCTGTGTGGGGGAGAAATGAACATG 840
Db 781 ACTGTCAACCAAGATGAATTAAGTGTGAACCCAGCTGTGTGGGGGAGAAATGAACATG 840
QY 841 CTCATTGAGCGGACTGTGGCTGGCACTGGCTGTCAGTGTCTCAGCATTTCCAAACA 900
Db 841 CTCATTGAACTACACTGGTGTGGAGTGGCTGTCTGCTGTCTCAGCATTTCCAAACT 900
QY 901 GTCTCTCCAGAAGTAAGAACGCTCTCATTTACTAGTGGGGGAAATGTAGACCTAA-- 958
Db 901 GTTTCCTCCAGAAGTAAGAACATTTGTTGCTCAGTGGTGAATGTAGACTTTAACC 960
QY 959 -CCTCCCTGACCTGGCTGGGAGGCTGAACGCCAGCTCTTACCAGAGCGTCTGTT 1016
Db 961 TCCTCCATACTTGGGTGAAGCAGGCTGAAAGCCAGCTCTTATCAGTGTGTTCTGT 1019
```

```
QY 841 CTCATTGACCGACTGCTGGCGTGACACGCTGGCTGAGTGTCTCTCAGCATTTCCTCAACA 900
|||||
Db 841 CTCATTGAACTTACAGCTGGTGTGGAGTGGCTGCTGCTCTCAACATTTTCAAACT 900
|||||
QY 901 GTCTCTCCAGAGTAAGAAGCTGTGCATTTGCTACAGTGGGGGGAATGTAGACCTAA-- 958
|||||
Db 901 GTTCTCCCAAGAGTAAGAAGCAATTTGTTATGTGCTCAGAGTGGTGGAAATGTAGACTTAACC 960
|||||
QY 959 -CCTCCCTCAACTGGGTGGGGGAGCTGAACGCCAGCTCCTTACCAGAGCTGCTGTTT 1016
|||||
Db 961 TCTCTCATTAACCTGGGTGAAGCAGGCTGAAGGCCAGCTTCTATCAGTCTGTTCTGT 1019
|||||

RESULT 5
AF169974 1166 bp mRNA linear PRI 27-OCT-2000
LOCUS Homo sapiens serine racemase mRNA, complete cds.
DEFINITION AF169974
ACCESSION AF169974.1 GI:11034784
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1166)
AUTHORS De Miranda,J., Santoro,A., Engelender,S. and Wolosker,H.
TITLE Human serine racemase: molecular cloning, genomic organization and
functional analysis
JOURNAL Gene 256 (1-2), 183-188 (2000)
MEDLINE 20510003
PUBMED 11034547
REFERENCE 2 (bases 1 to 1166)
AUTHORS DeMiranda,J., Santoro,A.H., Engelender,S. and Wolosker,H.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-1999) Departamento de Bioquímica Médica,
Instituto de Ciências Biomédicas, Universidade Federal do Rio de
Janeiro, Av. Brigadeiro Trompowski, s./n, Rio de Janeiro 21491-590,
Brazil

FEATURES
Source
1..1166
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue.type="brain"
17..1039
/function="catalyzes L- to D-serine racemization"
/codon_start=1
/product="serine racemase"
/protein_id="AAG27081.1"
/db_xref="GI:11034785"
/translation="MCAQYCSIFADVERKAHINRDSIHLTPVLTSILNQLTGRNLFF
KCELFOKTSFPIRGALNVRSLVDALRKPQAVVTHSSGNHGOALTVAALKLEGIPI
YIVVPOTAPDCKKLAIQAGSIVVCEPDESRENKRVETEGINVHPNQEPAVI
ACQGIATLEVLNQPLVDALVYPVGGGMLGAIATVRAKRSVKVYAEHPNSADQY
RKLKGLKPNLPYPPETADGQVSSIGLNTWPIIRDLDVDDIFVTDEIKCATQLWE
RNKLEIETAGVGAVALVSQHFQTSPEVKNICIVLSGGNVDLTSSITWVQAEKRPAS
YQSVSV"

BASE COUNT 333 a 238 c 275 g 320 t
ORIGIN

Query Match 79.4%; Score 807.8; DB 9; Length 1166;
Best Local Similarity 87.7%; Pred. No. 2.le-211;
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

QY 1 ATGTGTGCTCAGTACTGCATCTCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCAA 60
|||||
Db 17 ATGTGTGCTCAGTATTGCATCTCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCAA 76
|||||
QY 61 GACTCTATCCACTCACCCAGTGTCTAACAAAGCTCCATTTTGAATCAAAATAGCAGGCGC 120
|||||
Db 77 GATTCTATCCACTCACACGATGCTTAACAAAGCTCCATTTTGAATCAAAATAGCAGGCGC 136
|||||
QY 121 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAAACTGGGTCTTTTAAAGATTCCAGGTGCC 180
|||||
```

```
Db 137 AATCTTTTCTTCAATGTGAACCTCTTCCAGAAAAACAGGATCTTTTAAAGATTCTGTGCTCT 196
|||||
QY 181 CTTAATGCCATCAGAGGCTTAATTCCTGACACGCCAGAGAGAGCCCAAGCCGTAGTT 240
|||||
Db 197 CTTCAATGCCGTGAGAGCTTGGTTCTGATGCTTTTAGAAGAGGAGCCGAAGCTGTGTTT 256
|||||
QY 241 ACTCACAGCAGCGGAAACCATGCCAAGCTCTCACCTATATGCTGCTAACTTGAAGGAATT 300
|||||
Db 257 ACTCACAGCAGTGGAAACCATGCCAGGCTCTCACCTATGCTGCGCAAAATTTGAAGGAATT 316
|||||
QY 301 CTGCTTTACATTTGTTGTTCCCAACAGAGCTGCCAACTGCAAGAACTTGCACATTCACAGCC 360
|||||
Db 317 CCTGCTTATATTGTTGTTGCCCGCCAGACAGCTTCCAGAACTGTAAAAAACTTGCATACAGCC 376
|||||
QY 361 TATGAGCATCCATAGTATATCTGTACCCCAAGTGACGAGTCCAGAGAAAAAGTGCACCTCAA 420
|||||
Db 377 TAGGAGCGTCAATTTGTTACTGTGAACCTAGTAGTGAGTCCAGAGAAATGTGTCAGAAA 436
|||||
QY 421 AGAATTATCAAGAAACAGAGGAGCATTTTGGTCCATCCCAACCCAGAGGCTGCAGTGATA 480
|||||
Db 437 AGAGTTACAGAAACAGAGCATCATGGTACATCCCAACCCAGAGGCTGCAGTGATA 496
|||||
QY 481 GCTGACAAGGAACAATTCCTTGGAGTGTCTGGAACCGAGTTCCTTGGTAGATGACACTG 540
|||||
Db 497 GCTGACAAGGAGACAATTCCTTGGAGTGTCTGGAACCGAGTTCCTTGGTAGATGACACTG 556
|||||
QY 541 GTGTGTACCACTAGGAGGAGGAGGAATGTTGCTGGAATAGCCCATTAACAATTAGGCCCTG 600
|||||
Db 557 GTGTGTACCTGAGGTGGAGGAGGAGTGTGCTGGAATAGCAATTTACAGTTAAGGCTCTG 616
|||||
QY 601 AAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATCCAGATCCAGATGACTACCACT 660
|||||
Db 617 AAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAATGCCAGATGACTACCACTCC 676
|||||
QY 661 AAATGAAAGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGCTGTC 720
|||||
Db 677 AAGCTGAAGGGGAAACTGATGCCAACTCTTTATCTCTCCAGAAACCATAGCAGATGCTGTC 736
|||||
QY 721 AAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATTAAGAGACCTTCTGGATGATGCTTC 780
|||||
Db 737 AAATCCAGCATTTGGCTTGAACACCTTGGCCCTATTATTCAGGAGCCCTTGTGATGATATCTTC 796
|||||
QY 781 ACTGTCAACCAAGATGAAATCAAGTATGCAACCCAGCTGCTGTGGGGGAGAGTGAACACTG 840
|||||
Db 797 ACTGTCAACAGAGATGAAATTAAGTGTGCAACCCACCTGCTGTGGGAGAGATGAACACTA 856
|||||
QY 841 CTCATTGACCGACTGCTGGGTGGCACGTGGCTGACAGTGTGCTCAGCAGATTTTCCAAACA 900
|||||
Db 857 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT 916
|||||
QY 901 GTCTCTCCAGAGTAAGAAGCTGTCTCATTTACTCAGTGGGGGGAATGTAGACCTAA-- 958
|||||
Db 917 GTTCTCCCAAGAGTAAGAAGCAATTTGTTATGTGCTCAGTGGTGGAAATGTAGACTTAACC 976
|||||
QY 959 -CCTCCCTCAACTGGGTGGGGGAGGCTGAACGCCAGCTCCTTTACCAGAGCGTCTGTTT 1016
|||||
Db 977 TCTCTCATTAACCTGGGTGAAGCAGGCTGAAGGCCAGCTTCTTATCAGTCTGTTCTGT 1035
|||||

RESULT 6
AX224418 1770 bp DNA linear PAT 10-SEP-2001
LOCUS Sequence 1 from Patent WO0160987.
DEFINITION AX224418
ACCESSION AX224418
VERSION AX224418.1 GI:15554560
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1770)
AUTHORS Meyers,R.A. and Rudolph,Owen,L.A.
```

TITLE A human pyridoxal-phosphate dependent enzyme family member and uses therefor

JOURNAL Patent: WO 0160987-A 1 23-AUG-2001:  
Millennium Pharmaceuticals, Inc. (US)

FEATURES  
source Location/Qualifiers

1..1770  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
69..1091  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC69570.1"  
/db\_xref="GI:15554661"  
/translation="MCAQVCISFADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLF  
KCELFQKTSFKIRGALNVRSLVDALPKPAVTHSGNHGQALTYAAKLEGPIA  
YIYVQTPADCKKLATQAGSIVYCEPDESENVAKRVTETEGIMVHPNPEPAVI  
AGQGTALVELNVLVDALVPGVGGMLAGIAITKALKPSVKYIAEAPSNAADQY  
OSKLLKPLNLYPPETIADGVKSSIGLNTWPIRLDVIDFTVTEDEIKCATQLVWE  
RMKLLIPTAGVGAVALSHQFQTVSPVKNICIVLSGGNVDLTSSITWVQKAERPAS  
YQSVSV"

BASE COUNT 499 a 379 c 410 g 482 t  
ORIGIN

Query Match 79.4%; Score 807.8; DB 6; Length 1770;  
Best Local Similarity 87.7%; Pred. No. 2e-211;  
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;  
QY 1 ATGTGCTCAGTACTGCATCTCTCTTGTGTGATGTTGAAAGAGCTATATCAACATCAA 60  
DB 69 ATGTGCTCAGTATTGTCATCTCTTGTGTGATGTTGAAAGAGCTATATCAACATCGA 128  
QY 61 GACTCTATCCACCTCACCCAGTCTAAAGCTCCATTTTGAATCAATAGCAGGCGC 120  
DB 129 GATCTATCCACTCACACGAGTCTAACAGCTCCATTTTGAATCAATAGCAGGCGC 188  
QY 121 AATCTTTTCTCAAAATGTGAGCTCTTCCAGAAACTGGGTCCTTTTAAGATTCAGGTGCC 180  
DB 189 AATCTTTTCTCAAAATGTGAGCTCTTCCAGAAACTGGGTCCTTTTAAGATTCAGGTGCC 248  
QY 181 CTTAATGCCATCAGAGCTTAATCTCTGACACCCAGAGAGAGCCCAAGCCGTAGTT 240  
DB 249 CTCATGCGCTCAGAGCTTGGTCTCTGATGCTTTTAAAGAGGAGCGAAGCTGTGTT 308  
QY 241 ACTCACAGCAGCGAAGCACTGCGCAAGCTCTCACCTATGCTCTAACTGGAAGCAAT 300  
DB 309 ACTCACAGCAGTGAAGCACTGCGCAGGCTCTCACCTATGCTCTCAATTTGGAGCAAT 368  
QY 301 CTGCTTACATTTGTTTCCCAAGACAGTCCCACTGCAAGAACTGGCAATCCAGCC 360  
DB 369 CTGCTTATATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 428  
QY 361 TATGGAGCATCGATAGTATCTCTGACCAAGTGCAGAGTCCAGAGAAAGTCACTCA 420  
DB 429 TAGCGAGGTCGAATTTGATCTGTAACCTAGTGTGAGTCCAGAGAAATGTTGCAAA 488  
QY 421 AGAATTATCAAGAAACAGAGCATCTTGGTCCATCCCAAGCAGGAGCTGAGTGATA 480  
DB 489 AGAGTTACAGAAACAGAGGCATCATGGTACATCCCAAGCAGGAGCTGAGTGATA 548  
QY 481 GCTGGACAGGACAATTTGCCCTGGAAGTGTCTGAACCAAGTTCCTTTGGTAGTCACTG 540  
DB 549 GCTGGACAAGGACAATTTGCCCTGGAAGTGTCTGAACCAAGTTCCTTTGGTAGTCACTG 608  
QY 541 GTGGTACCAGTAGGAGGAGGAATGTTGCTGGAATAGCCATTAACAATTAGGCGCTG 600  
DB 609 GTGGTACCAGTAGGAGGAGGAATGTTGCTGGAATAGCAATTAAGTACAGGCTG 668  
QY 601 AAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCTACCACT 660  
DB 669 AAACCTAGTGTGAAGGTATATGCTGCTGAACCTCAATGCAGATGACTGCTACCACT 728  
QY 661 AAACCTGAAGGAACTGACCCCAATCTTCATCTCCAGAAACCATGACAGATGTGCT 720

DB 729 AAGCTGAAGGGGAAAGTATGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGC 788  
QY 721 AATCCAGCATTCGGCTTGAATACCTGGCCTATTATTAAGACACCTTGTGGATGATCTTTC 780  
DB 789 AATCCAGCATTCGGCTTGAACACCTGGCCTATTATCAGGAGCCTTGTGGATGATATCTTC 848  
QY 781 ACTGTCCACCAAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGGAGAGTGAACACTG 840  
DB 849 ACTGTCCACAGAGATGAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGATGAACACTA 908  
QY 841 CTCATTGAGCCGACTCTGCGGTGGCACTGGCTGCACTGCTGCTCAGCATTCAGATTTCCAAACA 900  
DB 909 CTCATTGAACCTACAGCTGGTGTGAGTGGCTGCTGCTGCTCAACATTTTCAAACT 968  
QY 901 GTCTCTCCAGAGTAAAGACGCTGCTCATTTGACTTACTCAGTGGGGGAGATGTAGACCTAA-- 958  
DB 969 GTTTCCTCCAGAGTAAAGAACATTTGATTTGCTCAGTGGTGGTGAATCTAGACTTTAAACC 1028  
QY 959 -CTCTCCTGAACCTGGGTGGGCGAGCTGAACGCCAGCTCCTTTACCAGACGGTCTGTTT 1016  
DB 1029 TCCTCCATTAACCTGGGTGAAGCAGCTGAAAGCCAGCTTCTTATCAGTCTGTTCTGT 1087

RESULT 7  
AK023169  
LOCUS Homo sapiens cDNA FLJ13107 fis, clone NT2RP3002501, weakly similar to THREONINE DEHYDRATASE CATABOLIC (EC 4.2.1.16).  
DEFINITION AK023169  
ACCESSION AK023169.1 GI:10434971  
VERSION  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens teratocarcinoma cell line: NT2 cDNA to mRNA, clone lib: NT2RP3 clone: NT2RP3002501.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Wagatsuma, M., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 2477)  
Isogai, T. and Otsuki, T.  
Direct Submission  
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

COMMENT  
FEATURES  
source Location/Qualifiers  
1..2477  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RP3002501"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP3"  
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction."  
69..1091  
/note="unnamed protein product"  
/codon\_start=1  
CDS

/protein\_id="BAB14442.1"  
/db\_xref="GI:10434972"  
/translation="MCAOVCISFADVERAHINIRDSIHLTPVLTSILNLTGRNLF  
KCELPQKTSFKIRGALANVRSLVDALERPKAVVTHSSGNHGQALYAAKLEGP  
YLVVPTAPDCKKLAIQAGSIVVCEPDESRENVRKYVEETEGIVHNPQEPAY  
AGOGTIALVNLQVPLDALVVPVGGGLAGIAITVAKRPSVKVYAAEENADDCY  
QSKLGLKMPNLYPETTADGVKSSIGLNTWPIIRDLDVDDIFTVTEDEIKATOLWE  
RKLLIEPTAGVGAVALSQQHFQTSPEVKNICIVLSGNNVDLTSSITWVQAE  
YQSVSV"  
BASE COUNT 738 a 535 c 524 g 680 t  
ORIGIN

Query Match 79.4%; Score 807.8; DB 9; Length 2477;  
Best Local Similarity 87.7%; Pred. No. 1.9e-211;  
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;  
QY 1 ATGTGTCCTCAGTACGCTCCCTTGTGATGTGAAAGAGCTCATATCAACATCAA 60  
DB 69 ATGTGTCCTCAGTACGCTCCCTTGTGATGTGAAAGAGCTCATATCAACATCAA 128  
QY 61 GACTCTATCCACCTCACCCAGTGTGTAAAGCTCATTTTGAATCAATAGCAGGCGC 120  
DB 129 GATTCTATCCACTCACACCTGCTGTACAGCTCCATTTGAATCAACTACAGGCGC 188  
QY 121 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAACTGGGTCTTTTAAAGATTGAGGTGCC 180  
DB 189 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAACTGGGTCTTTTAAAGATTGAGGTGCC 248  
QY 181 CTTAATGCCATCAGAGGCTTATTTCTGTACAGCCGACAGAGAAAGCCCAAGCGGTGTT 240  
DB 249 CTTAATGCCGTCAGAGGCTTGTCTGTGATGTGAAAGAGGAGCCGAAAGCTGTGTT 308  
QY 241 ACTCACAGCAGGAGAAACATGCGCAAGCTCTACCTATGCTGTCTTAACTGCAAGGAATT 300  
DB 309 ACTCACAGCAGTGGAAACCATGCGCAGGCTCTACCTATGCTGTCCCAATTTGGAAGGAAT 368  
QY 301 CTGCTTACATTTGGTGTCCCAACAGCTGCCAATGCTGTCTTAACTGCAAGGAATT 360  
DB 369 CTGCTTATTTGTTGTGTCGCCAGACAGCTCCAGAGCTGTAAAGAACTTGCATACAGCC 428  
QY 361 TATGAGCATCATAGTATGCTGACCAAGTGTGAGGCTCCAGAGAAAGTCACTCAA 420  
DB 429 TACGAGGCTCAATTTGATATCTGTGAACCTAGTGTGATGCTCCAGAGAAATTTGCAAAA 488  
QY 421 AGAATTTATGCAAGAAAGAGGATCTTGTGTCATCCCAACAGGAGGCTCAGTGATA 480  
DB 489 AGAGTTACAGAGAAACAGAGGATCATGTGTACATCCCAACAGGAGGCTCAGTGATA 548  
QY 481 GCTGACAGAGAACTGTCCTGGAAGTGTGACAGGTTCCCTTGGTAGATGCACTG 540  
DB 549 GCTGACAGAGGAGCAATTTGCCCTGGAAGTGTGACAGGTTCCCTTGGTAGATGCACTG 608  
QY 541 GTGGTACAGTAGGAGGAGGAGGAATGTTGTGGAATAGCCATACAAATTAAGGCCCTG 600  
DB 609 GTGGTACCTGTAGTGGAGGAGGAATGTTGTGGAATAGCAATTTACAGTTAAGGCTCG 668  
QY 601 AAACCTAGTGTGAAGGTATACGCTGTGAGCCCTCGAATGCAAGTCACTGTACCAGTCT 660  
DB 669 AAACCTAGTGTGAAGGTATATGCTGTGAACCTCAATGCAAGTCACTGTACCAGTCT 728  
QY 661 AAACCTAGGAGGAACTGACCCCAATCTTATCTCCAGAAACCATAGCAGATGGTGTG 720  
DB 729 AAGCTGAGGAGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTG 788  
QY 721 AAATCCAGCATGGCTGGAATACCTGGCTTATTAAGAGACCTTTGTGATCATCTCTC 780  
DB 789 AAATCCAGCATGGCTGGAACACCTTGTGATCATCTCTTATCAGGAGCTTTGTGATCATCTC 848  
QY 781 ACTGTCACCAAGATCAATCAAGTATGCAACCCAGCTGGTGTGGGGAGGAATGAACCTG 840  
DB 849 ACTGTCACAGAGGATGAATTAAGTGTGCAACCCAGCTGGTGTGGGAGGATGAACCTA 908  
QY 841 CTCATTGAGCCGACTGTGCGGTGGCACTGGGTGCTGCTGTCTCAGCATTTTCCAAACA 900

DB 909 CTCAATTGAACCTACAGCTGGTGTGGAGTGGCTGTCTCTCAACATTTTCAAACT 968  
QY 901 GTCTCTCCAGAGTAAAGAACCTCTGCATTTACTACAGTGGGGGAGTATAGACCTAA-- 958  
DB 969 GTTTCCTCCAGAGTAAAGAACATTTGTATGTGCTCAGTGGTGGAAATGTAGACTTAACC 1028  
QY 959 -CCTCCCTGAACCTGGGTGGGCGAGCTGAACGCCAGCTCTTACCAGACGGTCTCTTT 1016  
DB 1029 TCCTCCATAACTTGGGTGAAGCAGGCTGAAGCCAGCTTCTTATCAGTCTGTTCTGT 1087  
RESULT 8  
AX259575 LOCUS AX259575 1336 bp DNA linear PAT 26-OCT-2001  
DEFINITION Sequence 1 from Patent WO0173077.  
ACCESSION AX259575  
VERSION AX259575.1 GI:16508702  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE 1  
AUTHORS Ramakrishnan,S.  
TITLE Regulation of human serine racemase enzyme  
JOURNAL Patent: WO 0173077-A 1 04-OCT-2001;  
Bayer Aktiengesellschaft (DE)  
FEATURES  
source location/Organism  
1. 1336  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 386 a 268 c 319 g 362 t 1 others  
ORIGIN  
Query Match 79.2%; Score 806.2; DB 6; Length 1336;  
Best Local Similarity 87.6%; Pred. No. 5.7e-211;  
Matches 893; Conservative 0; Mismatches 123; Indels 3; Gaps 1;  
QY 1 ATGTGTCCTCAGTACGCTCCCTTGTGATGTGAAAGAGCTCATATCAACATTCAA 60  
DB 52 ATGTGTCCTCAGTACGCTCCCTTGTGATGTGAAAGAGCTCATATCAACATTCGA 111  
QY 61 GACTCTATCCACCTCACCCAGTGTGTAAAGCTCCATTTTGAATCAATAGCAGGCGC 120  
DB 112 GATTCTATCCACTCACACCTGCTGTACAGCTCCATTTTGAATCAACTAACAGGCGC 171  
QY 121 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAACTGGGTCTTTTAAAGATTGAGGTGCC 180  
DB 172 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAACAGGATCTTTTAAAGATTGAGGTGCT 231  
QY 181 CTTAATGCCATCAGAGGCTTAATTTCTGACAGCCGACAGAGAGCCCAAGCCGTAGTT 240  
DB 232 CTCATAGCGCTCAGAGCTTGGTCTCTGATGCTTTTGAAGAGGAGCCGAAAGCTGTT 291  
QY 241 ACTCACAGCAGGAGAAACCATGCGCAAGCTCTCACCTATGCTGTCTTAACTGGAAGGAAT 300  
DB 292 ACTCACAGCAGTGGAAACCATGCGCAGGCTCTCACCTATGCTGCCAAATTTGAAGGAAT 351  
QY 301 CCGTCTTACATTTGGTGTCCCAACAGCTCCCACTGCAAGAACTGCAATTCAGGCC 360  
DB 352 CCGTCTTATTTGTTGGTGTCCCAACAGCTCCAGAGCTGTAAAGAACTTCAATACAGCC 411  
QY 361 TATGAGCATTCGATAGTACTGTGACCAAGTGTGAGCTCCAGAGAAAGTCACTCAA 420  
DB 412 TACGAGGCTCAATTTGTATCTGTGAACCTAGTGTGATGAGTCCAGAGAAATTTGCAAAA 471  
QY 421 AGAATTTATGCAAGAAACAGGCTATTTGGTCTTCCCAACAGGAGGCTCAGTGATA 480  
DB 472 AGAGTTACAGAGAAACAGAGGCTATGTTGATACATCCCAACAGGAGGCTCAGTGATA 531  
QY 481 GCTGACAGGAGAACTTGGCTTGGAGTGTGAAAGTGTGAAAGGTTCCCTTGGTAGATGCTG 540

```
Db 532 GCTGGACAAGGGAATTCCTCCCTGGAAGTCTGTAACCAAGTCTCTTTGGTGGATGCAGT 591
Qy 541 GTGCTACCACTAGGAGGAGGAGGAATGTTGCTGGAATAGCCATTACAATTAAGGCCCTG 600
Db 592 GTGGTACCTGTAGTGGAGGAGGAATGCTTGTGGAATACCAATACAGTTACAGCTCTG 651
Qy 601 AAACCTAGTGTGAAGGTATACGCTGTGAGCCCTGCAATCAGATCAGATCCTACCACTCT 660
Db 652 AAACCTAGTGTGAAGGTATATGCTGTGGAACCTCAATCAGATCAGATCCTACCACTCT 711
Qy 661 AAACCTAGTGTGAAGGTATACGCTGTGAGCCCTGCAATCAGATCAGATCCTACCACTCT 720
Db 712 AAGCTGAAGGGGAAACTGATGCGCAATCTTTATCTCCAGAAACCATAGCAGATGGTCT 771
Qy 721 AAATCCAGCATTCGCTTGAATACCTGCGCTTATTAAGAGACCTTGTGGATGATGCTTC 780
Db 772 AAATCCAGCATTCGCTTGAACACCTGCGCTTATTAAGAGACCTTGTGGATGATGCTTC 831
Qy 781 ACTGTCCACCAAGATGAATCAAGTATGCAACCCAGCTGGTGGGGGAGAAATGAACCTG 840
Db 832 ACTGTCCACAGAGTGAATTAAGTGTGCAACCCAGCTGGTGGGGAGAGATGAACCTA 891
Qy 841 CTATTCAGCCGACTGCTGCGTGGCACTGGCTGAGTGTGCTGCTCAGCATTTCCAAACA 900
Db 892 CTATTCAGCCGACTGCTGCGTGGCACTGGCTGAGTGTGCTGCTCAGCATTTCCAAACT 951
Qy 901 GTCTCTCCAGAAGTAAAGACGCTGCTATGTTACTCAGTGGGGGGAATGTAGACCTAA-- 958
Db 952 GTTTCCTCCAGAAGTAAAGACATTTGTTATGCTGCTCAGTGGTGGAAATGTAGACTTAA 1011
Qy 959 -CTCCCTCAACTGGTGGGGGAGGCTGAGTGAACCCAGCTCCTTACCAGCGGCTCTGTTT 1016
Db 1012 TCCTCCCACTAATTTGGGTGAAGCAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGT 1070

RESULT 9
HSM805450
LOCUS Homo sapiens mRNA; cDNA DKFp762A2415 (from clone DKFp762A2415).
DEFINITION AL834378
ACCESSION AL834378.1 GI:21740048
VERSION 1
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2264)
AUTHORS Ansoorge,W., Wirkner,U., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
COMMENT Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFp762A2415) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/.
Location/Qualifiers
1..2264
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="22.9 cR from top of Chr17 linkage group"
/clone="DKFp762A2415"
/tissue_type="melanoma (Mewo cell line)"
/clone_lib="762 (synonym: hmel2). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="adult"
1..2264
/gene="DKFp762A2415"
<1..867

gene
CDS
```

```
/gene="DKFp762A2415"
/notes="similarity to serine racemase (Mus musculus)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD39041.1"
/db_xref="GI:21740049"
/translation="GSFKIRGALNVRSLVDPALERKRAVTVHSSGNHQAALYAAK
LEGIPAYIVVPTAPDKKLAIOAGSIVYCPSPDSRENAKRVTEETEGIMVHPN
QBPVATAGGTALVNLQVPLVDALVPGVGGMLAGTAITVKALKPSVKVYAAPS
NADDCYQSKLKGKLPNLIPPTIAGVYKSSIGTSVNPRIIRLDVDDIFVTVEDEIKCA
TOLVBERMKLLIEPTAGVGVAAVLVSHQFOTVSPVKNKICIVLGGNVDLTSSITWVKQ
AERPASYQSVSV
2226..2231
polyA_signal
polyA_site
2253
/gene="DKFp762A2415"
BASE COUNT 692 a 478 c 465 g 629 t
ORIGIN
Query Match 65.3%; Score 664.6; DB 9; Length 2264;
Best Local Similarity 86.4%; Pred. No. 5.5e-172;
Matches 746; Conservative 0; Mismatches 114; Indels 3; Gaps 1;
Qy 157 GGTCTTTTAAAGATTCGAGGTGCCCTTAATGCCATCAGAGGCTTAATTCCTGACACGCCA 216
Db 1 GGATCTTTTAAAGATTCGTGTGCTCTCAATGCCGTCAGAAGCTTGGTCTGATGCTTTA 60
Qy 217 GAAGAGAAGCCCAAGCCGTAGTTACTCAGCAGAGCGGAAACCATGCCAGCTCTCACC 276
Db 61 GAAAGGAAGCCGAAGAGCTGTTGTACTCAGCAGGTGGAAACCATGCCAGGCTCTCACC 120
Qy 277 TATGCTGCTAACTGGAAGGAATTCCTGCTTACATTTGTTGTTCCCAAAACAGCTCCCAAC 336
Db 121 TATGCTGCCAATTTGAAGGAATTCCTGCTTATTTGTTGTCGCCACAGACAGCTCCAGAC 180
Qy 337 TGAAGAAACTGCAATCAAGCTATGGAGCATCGATAGTATATCTGTGACCCCAAGTGAC 396
Db 181 TGTAAAAAATTGCAATACAGGCTACGGAGCGCTCAATTTGATCTGTGAACCTAGTGAT 240
Qy 397 GAGTCCAGAGAAAGTCACTCAAGAAATTTCCAGAAACACAGAAAGCATCTTGGTCCAT 456
Db 241 GAGTCCAGAGAAATTTGCAAAAGAGTTACAGAAAGAAACAGAAAGCATCTGGTACAT 300
Qy 457 CCCAACGAGGAGCTGCAGTATAGCTGGACAAGAAACAATTCCTCGGAAGTGCCTGAAC 516
Db 301 CCCAACGAGGAGCTGCAGTATAGCTGGACAAGGACAAATTCCTCGGAAGTGCCTGAAC 360
Qy 517 CAGGTTCCCTTGGTGTAGTGCATCTGGTGGTACCATAGTAGGAGGAGGAATGGTTGTGGA 576
Db 361 CAGGTTCCCTTGGTGTAGTGCATCTGGTGGTACCTGTAGTGGAGGAGGAATGGTTGTGGA 420
Qy 577 ATAGCATTACAAATTAAGCCCTGAAACCTAGTGTGAAGGTATACGCTGTGAGCCCTCG 636
Db 421 ATAGCAATTTACAGTTAAGGCTCTGAAACCTAGTGTGAAGGTATATGCTGCTGAACCTCA 480
Qy 637 AATGCAGATGACTGCTACCACTGCTAAACTGAAAGAGAGAACTGACCCCAATCTTCATCT 696
Db 481 AATGCAGATGACTGCTACCACTGCTGAAAGGGAACAACTGATGCCCAATCTTTATCTCT 540
Qy 697 CCAGAAACCATAGCAGATGGTGTCAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATA 756
Db 541 CCAGAAACCATAGCAGATGGTGTCAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATC 600
Qy 757 AGAGACCTTGTGGATGATGCTTTCATCTGTCACCAAGAGATCAATCAAGTATGCAACCCAG 816
Db 601 AGGACCTTGTGGATGATGATCTTTCATCTGTCACAGAGGATGAAATTAAGTGTGCAACCCAG 660
Qy 817 CTGGTGTGGGGAGAATGAACTGCTCATTGAGCCGACTGCTGGCTGGCAGCTGGCTGCA 876
Db 661 CTGGTGTGGGAGAGGATGAACTGCTCATTGAACTTACAGCTGCTGTTGGAGTGGCTGCT 720
Qy 877 GTGCTCTCAGCATTTCCAAACAGCTCTCTCCAGAAAGTAAAGAACGCTGTGCAATGTACTC 936
|||||
```

```

Db      721 GTGCTGCTCAACATTTTCAAACTGTTTCCCGAGAGTAAAGAACATTTGTATTGTGCTC 780
QY      937 AGTGGGGGAATGTAGACATAA---CCTCCCTGAACCTGGGTGGGCGAGGCTGAACGGCCA 993
Db      781 AGTGGTGAATGTAGACTTACCTCTCCATAACTTGGGTGAACGAGGCTGAAGGGCCA 840
QY      994 GTCCTCTACACGAGCGGTCTGTTT 1016
Db      841 GCTTCTTATCAGTCTGTTTCTGT 863

RESULT 10
AL604066/c
LOCUS
DEFINITION
AL604066 184151 bp DNA linear ROD 14-MAR-2002
complete sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 184151)
Direct Submission
Submitted (14-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
Clone requests: clonerequests@sanger.ac.uk
On Mar 21, 2002 this sequence version replaced gi:17902958.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
RP23-174M12 is
from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
This sequence is the entire insert of clone RP23-174M12. The true
right end of clone RP23-194P5 is at 82810 in this sequence.

FEATURES
source
1. 184151
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-174M12"
/clone_lib="RPCI-23"
BASE COUNT 46879 a 41019 c 43482 g 52771 t
ORIGIN
Query Match 25.3%; Score 257.2; DB 10; Length 184151;
Best Local Similarity 73.1%; Pred. No. 6.1e-60;
Matches 421; Conservative 0; Mismatches 3; Indels 152; Gaps 1;
QY 593 AGGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT 652
Db 125770 AGGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT 125711
QY 653 ACCAGTCTAACTGAAGAGAGAACTGACCCCAATCTTCATCCTCCAGAAACCATAGCAG 712

```

```

Db      125710 ACCAGTCTAACTGAAGAGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAG 125651
QY      713 ATGGTGTCAAAATCCAGCATTTGCTTGAATACCTGGCCCTATTATAAGAGACCTTTGGTGATG 772
Db      125650 ATGGGTGTCAAAATCCAGCATTTGCTTGAATACCTGGCCCTATTATAAGAGACCTTTGGTGATG 125591
QY      773 ATGCTCTTCACTGTCAACCCGAAGATGAATCA----- 802
Db      125590 ATGCTCTTCACTGTCAACCCGAAGATGAATCAAGGTAGACTCAGGCATTGTTAACAGAAC 125531
QY      803 ----- 802
Db      125530 CTGACAGTAGATCACATACTTATCACAACAAACCCAAATGTCTACTTCTATTTTTCGAATG 125471
QY      803 ----- 802
Db      125470 TAACTACTAGACTCCGGCTCCGTTCTTTGTTTCTATTAAACCCCTTGTCTCTTTCTGTGGCCA 125411
QY      803 --ATGATGCAACCCAGCTGGTGTGGGGAGAGAAATGAAACTGCCTATTGAGCCGACTGCTGG 860
Db      125410 ACAGTATGCAACCCAGCTGGTGTGGGGAGAGAAATGAAACTGCTCATTGACCCGACTGCTGG 125351
QY      861 COTGGCACTGGTGTGAGTGTCTCTCAGCATTTTCCAAACAGTCTCTCCAGAAAGTAAAGAA 920
Db      125350 COTGGCACTGGTGTGAGTGTCTCTCAGCATTTTCCAAACAGTCTCTCCAGAAAGTAAAGAA 125291
QY      921 COTGTGCTATTGTACTGACGTGGGGGGAATGTACACCTAACTCCCTGAACTGGTGGGGCA 980
Db      125290 COTGTGCTATTGTACTGACGTGGGGGGAATGTACACCTAACTCCCTGAACTGGTGGGGCA 125231
QY      981 GGCTGAAACGGCCAGCTCTTACACACGCGTCTGTTT 1016
Db      125230 GGCTGAAACGGCCAGCTCTTACACACGCGTCTGTTT 125195

RESULT 11
AC044847/c
LOCUS
DEFINITION
AC044847 205910 bp DNA linear HTG 07-DEC-2000
SEQUENCE, 37 unordered pieces.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 205910)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Melidrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J.,

```



```
misc_feature /note="assembly_fragment"
21050..23372
misc_feature /note="assembly_fragment"
23473..26193
misc_feature /note="assembly_fragment"
26294..28324
misc_feature /note="assembly_fragment"
28425..31034
misc_feature /note="assembly_fragment"
31135..34444
misc_feature /note="assembly_fragment"
34545..38917
misc_feature /note="assembly_fragment"
39018..43628
misc_feature /note="assembly_fragment"
43729..47588
misc_feature /note="assembly_fragment"
47689..52019
misc_feature /note="assembly_fragment"

Query Match 25.3%; Score 257.2; DB 2; Length 205910;
Best Local Similarity 73.1%; Pred. No. 6e-60;
Matches 421; Conservative 0; Mismatches 3; Indels 152; Gaps 1;

QY 593 AGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT 652
Db 147884 AGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT 147825

QY 653 ACCAGTCTAACTGAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAG 712
Db 147824 ACCAGTCTAACTGAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAG 147765

QY 713 ATGGTGTCAATCCAGCATTTGGCTGAATACCTGGCTATTATAGAGACCTTTGGATG 772
Db 147764 ATGGTGTCAATCCAGCATTTGGCTGAATACCTGGCTATTATAGAGACCTTTGGATG 147705

QY 773 ATGCTCTACTGTACCCGGAAGATGAATCA-----802
Db 147704 ATGCTCTACTGTACCCGGAAGATGAATCAAGTGAAGTCAAGGCTAGGCAATTGTTAACAGAAC 147645

QY 803 -----802
Db 147644 CTGACAGTAGATCACATATCTATACAAACCCCAATGTCTACTTCAATTTTCAAGATG 147585

QY 803 -----802
Db 147584 TAACCTACTAGACTCGGCTCGTTTCTTTTCATTAAACCCTTGCTCTCTTCTGTGCCA 147525

QY 803 ---AGTATGAACCCAGCTGGTGTGGGGGAGATGAACCTGCTCAATGAGCCGACTGCTGG 860
Db 147524 ACAGTATGAACCCAGCTGGTGTGGGGGAGATGAACCTGCTCAATGAGCCGACTGCTGG 147465

QY 861 CGTGGCAGTGGTGCAGTCTGCTCAGCATTTCCAAACAGTCTCTCCAGAGTAAGAA 920
Db 147464 CGTGGCAGTGGTGCAGTCTGCTCAGCATTTCCAAACAGTCTCTCCAGAGTAAGAA 147405

QY 921 CGTCTCATTTGACTCAGTGGGGGGAATGTAGACCTTAACCTCCCTCAACTGGGTGGGGCA 980
Db 147404 CGTCTCATTTGACTCAGTGGGGGGAATGTAGACCTTAACCTCCCTCAACTGGGTGGGGCA 147345

QY 981 GGTGTAACGCCAGCTCTTACACAGCGGTCTGT 1016
Db 147344 GGTGTAACGCCAGCTCTTACACAGCGGTCTGT 147309
```

```
RESULT 12
BC031531/c 3379 bp mRNA linear ROD 07-AUG-2002
LOCUS Mus musculus, Similar to hypothetical protein FLJ10534, clone
DEFINITION MGC:28320 IMAGE:4014819, mRNA, complete cds.
ACCESSION BC031531
VERSION BC031531.1 GI:21619397
KEYWORDS MGC.
```

```
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3379)
REFERENCE Strausberg,R.
AUTHORS Direct Submission
TITLE National Institutes of Health, Mammalian
JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabp@remail.nih.gov
Tissue procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: ang@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 36 Row: m Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.
FEATURES
Location/Qualifiers
1..3379
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C2ECH II"
/clone="MGC:28320 IMAGE:4014819"
/tissue_type="Mammary tumor metastasized to lung.
MTV-LTR/Wnt1 model. Expression driven by an MTV-LTR
enhancer."
/clone_lib="NCI_CGAP_Lu30"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
457..2187
/codon_start=1
/product="Similar to hypothetical protein FLJ10534"
/protein_id="AAH31531.1"
/db_xref="GI:21619398"
/translation="MAKVAADTILFLDPLEGWDSTGDYCLSCFLAOGLPYTLAVOGL
SGFPKKQIDARKKLSKTVKRPEDKLLLDTOQESGMLRLANOKORHLAFDRRR
AVLFAHADFVPESESDLVGTLKISGYVRGRLTNVNSLLHYVGHGFQNDADVPDP
FLNPRVTKSQKPNMAVEVCVTDAAADMEEDLVLMKADPDHQSLSQTEAIPDPMEG
EFTWTEELDADLLKRSRVKVPKTSYQAEWILDEGSDGEGGYDDIOH
EGFMEESQDGGSEEEETMTLGESVRDDLDYDEKVEDDEERMLERYKQERLEEM
FPDEMTPRDVAARIFQRYGLKSPRTSPWDPKENVLPDRYARIFQFQNFVTKRIF
KTEKEAEGAEVGVVTLHVSQVPSVVEYFROGAPLIAFSLLPYEOKMSVLNVVS
RNPQTEPVKAEKEELIFHCGFRFRASPLFSOHTADKKHFORFLTADALVTVFAP
ITFPASVLLFKQRRNGMSLIATGHLSFVDPDRWVKRVVLSGHPFKFTKMAVVRY
MFFNRGCDVV"
BASE COUNT 958 a 694 c 887 g 840 t
ORIGIN
Query Match 24.6%; Score 250.8; DB 10; Length 3379;
Best Local Similarity 72.4%; Pred. No. 5.3e-58;
Matches 417; Conservative 0; Mismatches 7; Indels 152; Gaps 1;
QY 593 AGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT 652
Db 3223 AGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT 3164
QY 653 ACCAGTCTAACTGAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAG 712
```

Db 3163 ACCAGTCTAACTGAAGAGAACTGACCCCAATCTTCTCCAGAAACCATGACG 3104

QY 713 ATGGTGCAATCCAGCATGGCTGTAATACCTGGCTATTATAGAGACCTTGTGGATG 772

Db 3103 ATGGTGCAATCCAGCATGGCTGTAATACCTGGCTATTATAGAGACCTTGTGGATG 3044

QY 773 ATGCTTCACTGTCCAGCAAGATGAATCA----- 802

Db 3043 ATGCTTCACTGTCCAGCAAGATGAATCAAGCTAGACTCAGGCATTGTTAACAGAAC 2984

QY 803 ----- 802

Db 2983 CTGACAGTAGATCACATACCTATTATACAAACCAACCAAAATGCTACCTCAATTTTCGAATG 2924

QY 803 ----- 802

Db 2923 TAACCTACTAGACTCGGCTCCGTTCTTGTTTCATTAAACCTTGTCTCTTCTGTGCCA 2864

QY 803 --AGTATGCAACCCAGCTGGTGTGGGGAGAAATGAACCTGCTCATTCAGCCGACTGCTGG 860

Db 2863 ACAGTGTGCAACCCAGCTGGTGTGGGGAGAAATGAAGCTGCTCATTCAGCCGACTGCTGG 2804

QY 861 CGTGGCACTGGCTGCTGCTCAGCATTTCCAAACAGTCTCTCCAGAACTAAGAA 920

Db 2803 CGTGGCACTGGCTGCTGCTCAGCATTTCCAAACAGTCTCTCCAGAACTAAGAA 2744

QY 921 CGTCTGCTACTGCTCAGTGGGGGAGATGAGACCTAACCTCCCTGAACCTGGGTGGGCA 980

Db 2743 CGTCTGCTACTGCTCAGTGGGGGAGATGAGACCTAACCTCCCTGAACCTGGGTGGGCA 2684

QY 981 GCGTGAACGCCAGCTCCTTACCAGACGGTCTCTGTT 1016

Db 2683 GCGTGAACGCCAGCTCCTTACCAGACGGTCTCTGTT 2648

## RESULT 13

AC122087/c  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-280H23, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 61 unordered pieces.  
AC122087 101814 bp DNA linear HTG 23-JUL-2002  
AC122087.2 GI:21903132  
VERSION HTG: HTGS\_PHASE1.  
KEYWORDS Rattus norvegicus.  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 101814)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaral-Tunget,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Haves,A., Hernandez,J., Hernandez,O., Hodson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratochvic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tameris,K., Tang,H., Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G., and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 101814)  
Worley,K.C.  
Direct Submission  
Submitted (22-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 101814)  
Worley,K.C.  
Direct Submission  
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 18, 2002 this sequence version replaced gi:21070489.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GW21  
Center clone name: CH230-280H23  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 43731 bases at least Q40  
Consensus quality: 47905 bases at least Q30  
Consensus quality: 51501 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 61 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 1045: contig of 1045 bp in length  
\* 1046 1145: gap of unknown length  
\* 1146 2200: contig of 1055 bp in length  
\* 2301 3345: gap of unknown length  
\* 3346 3445: contig of 1045 bp in length  
\* 3446 4445: gap of unknown length  
\* 4446 4545: contig of 1000 bp in length  
\* 4546 5575: gap of unknown length  
\* 5576 5675: contig of 1021 bp in length  
\* 5676 6697: gap of unknown length  
\* 6697 7807: contig of 1011 bp in length  
\* 7808 7907: gap of unknown length  
\* 7908 8910: contig of 1003 bp in length  
\* 8911 9010: gap of unknown length



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 3807)  
Strausberg, R.  
Direct Submission  
Submitted (07-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapsb-re@mail.nih.gov](mailto:cgapsb-re@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nih.gov](mailto:nisc_mgc@nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J., Tjongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL plate: 39 Row: j Column: 10  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8922495.

## FEATURES

Source

Location/Qualifiers  
1..3807  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:29514 IMAGE:4896567"  
/tissue\_type="Brain, neuroblastoma"  
/clone\_lib="NIH MGC\_19"  
/lab\_host="PH10B-R"  
/note="Vector: pOTB7"  
319..2289  
/codon\_start=1  
/product="Unknown (protein for MGC:29514)"  
/protein\_id="AAH19090.1"  
/db\_xref="GI:17512218"  
/translation="MAKVADTILFLDPLEGWDSTGDCVLSCLFAGQLPTTYTFLAVQGI  
SGLPKKQIDTRKLSKAVEKRPDKLKLDDTQOEAGMLRLQLANQOHLAFDRDR  
AYLFAHVDVFPSENNLVGTLKISGYRGOTLVNRLHIVYGDQFMQKIDAPDGP  
FPLNPRGKPKQDPDAMEICATDAVDOMEGLKVMKADPGRQESLOAEVTPDPMEG  
EOTWTEELSEAKDFLKEGSKVKKVPKTSYOAENWLDGSGSGGDEYEVDDM  
EHEDMEESQDESEEEYEYTWIGESVHDDLDYKVDDEAEAKMLEKYQERLEE  
MPDEVDTPDVAARIRQKYRGLKSFRTSPWDPKNELPQDYARIFQFONFTNRKSI  
FRKEVEKEVEAGVYTLHVSEVPSVVEFCRQGTPLIAFSLPHQKMSLVNMY  
RRDPNGTEVKAKEELIFHCGRFRFRASPLFSQHTAADKHLQRELTADMLAVIYA  
PTTFPASVLLFKVQKSNHMSLIATGHLMSVDPDRMWTKRVVLSGHPKFIETKMAVVR  
YMFENREDVLFKPVLEKTRKGRGRHKEPLGLTHGHMKCSFDGKLKSQDTVLNLYKR  
VFPKTYDYPYEPVPWLKSEISSVPOGGME"

## CDS

319..2289

## BASE COUNT

ORIGIN

1102 a 813 c 943 g 949 t  
Query Match 17.9%; Score 181.8; DB 9; Length 3807;  
Best Local Similarity 88.0%; Pred. No. 5.3e-39;  
Matches 198; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
QY 593 AGGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT 652  
|||||  
DB 3351 AGGCTCTGAACCTAGTGTGAAGGTATATGCTGCTGAACCTCAACCTGAGATGACTGCT 3292  
|||||  
QY 653 ACCAGCTCTAACTGAAGAGGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAG 712  
|||||  
DB 3291 ACCAGTCCAAGCTGAAGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAG 3232  
|||||

QY 713 ATGCTGTCAAAATCCACGATTGGCTTCAATACCTGGCTATTATATAGAGACCTTGTGGATG 772  
|||||  
DB 3231 ATGCTGTCAAAATCCACGATTGGCTTCAACACCTGGCTATTATATAGAGGACCTTGTGGATG 3172  
|||||  
QY 773 ATGCTGTCACTGTCACCGAAGATGAATCAAGTATGCAACCCAGC 817  
|||||  
DB 3171 ATATCTTCACTGTCACAGAGGATGAATTAAGGTAGAGCTCCAGC 3127  
|||||  
RESULT 15  
AK026565/c 4140 bp mRNA linear PRI 29-SEP-2000  
LOCUS  
DEFINITION Homo sapiens cDNA: FLJ22912 fis, clone KAT06132.  
ACCESSION AK026565  
VERSION AK026565.1 GI:10439447  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens signet-ring cell carcinoma cell\_line:KATO III cDNA to mRNA, clone\_lib:KAT clone:KAT06132.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Watanabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.  
NEDO human cDNA sequencing project  
TITLE Unpublished  
JOURNAL 2 (bases 1 to 4140)  
REFERENCE Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.  
Direct Submission  
TITLE Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: [cdna@ims.u-tokyo.ac.jp](mailto:cdna@ims.u-tokyo.ac.jp), Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).  
FEATURES Location/Qualifiers  
Source 1..4140  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="KAT06132"  
/cell\_line="KATO III"  
/cell\_type="signet-ring cell carcinoma"  
/clone\_lib="KAT"  
/note="cloning vector pME18SFL3"  
BASE COUNT 1195 a 901 c 1050 g 994 t  
ORIGIN  
Query Match 17.9%; Score 181.8; DB 9; Length 4140;  
Best Local Similarity 88.0%; Pred. No. 5.3e-39;  
Matches 198; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
QY 593 AGGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT 652  
|||||  
DB 3507 AGGCTCTGAACCTAGTGTGAAGGTATATGCTGCTGAACCTCAACCTGAGATGACTGCT 3448  
|||||  
QY 653 ACCAGCTCTAACTGAAGAGGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAG 712  
|||||  
DB 3447 ACCAGTCCAAGCTGAAGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAG 3388  
|||||  
QY 713 ATGCTGTCAAAATCCACGATTGGCTTGAATACCTGCGCTATTATATAGAGACCTTGTGGATG 772  
|||||  
DB 3387 ATGCTGTCAAAATCCACGATTGGCTTGAACACCTGGCTTATTATAGAGGACCTTGTGGATG 3328  
|||||  
QY 773 ATGCTTCACTGTCAACCGAAGATGAATCAAGTATGCAACCCAGC 817

Db 3327 ATATCTCACTGTCACAGAGGATGAAATTAAGGTGAGGCTCCAGC 3283  
|||||  
|||||

Search completed: June 24, 2003, 05:10:53  
Job time : 2676 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 03:25:53 ; Search time 286 Seconds  
(without alignments)  
8015.854 Million cell updates/sec

Title: US-09-889-609B-1  
Perfect score: 1018  
Sequence: 1 atgtgtcactgactgcat.....ttaccagacgtgtgttaa 1018

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1018	100.0	1018	21	AAA59294
2	1011.2	99.3	1672	21	AAA59300
3	807.8	79.4	1023	21	AAA59299
4	807.8	79.4	1023	22	AAS15217
5	807.8	79.4	1770	22	AAD14461
6	807.8	79.4	2477	22	AAH16282
7	806.2	79.2	1336	22	AAI70575
8	806.2	79.2	2674	22	AAH165020
9	599.6	58.9	848	22	AAH06600

10	530.4	52.1	731	24	ABL90122	Human polynucleoti
11	434.2	42.7	608	21	AAA59295	N-terminal sequenc
12	240.8	23.7	509	21	AAA59296	C-terminal sequenc
13	181.8	17.9	20892	22	ABAL5709	Human nervous syst
14	168.2	16.5	861	21	ACGL0869	Human secreted pro
15	139.8	13.7	31096	18	AAV74370	Staphylococcus aur
16	130	12.8	583	22	AAH98041	Murine 7-transmemb
17	102	10.0	3550	23	AAH98041	DNA encoding novel
18	90.2	8.9	1189	23	ABL18395	Drosophila melanog
19	90.2	8.9	1447	23	ABL13783	Drosophila melanog
20	87.4	8.6	1830121	17	AAT42063	Haemophilus influe
21	84.8	8.3	1866	23	AAH82540	DNA encoding novel
22	84.8	8.3	1866	23	AAH82540	DNA encoding novel
23	77	7.6	954	24	ABK77458	Bacillus clausii g
24	75.4	7.4	542	22	AAH83752	Human polynucleoti
25	71.6	7.0	3402	23	ABL18394	Drosophila melanog
26	71.6	7.0	4179	23	ABL13782	Drosophila melanog
27	71	7.0	105184	24	ABK24122	Bacterial artifi
28	70.2	6.9	930	22	AAH66056	C glutamicum codin
29	70.2	6.9	1053	22	AAF71822	Corynebacterium gl
30	70.2	6.9	349980	22	AAH68527	C glutamicum codin
31	69.2	6.8	922	18	AAT84000	DNA encoding a thr
32	67.2	6.6	1478	20	AAH07180	Corn threonine dea
33	66.4	6.5	1509	20	AAH21199	Mutant threonine d
34	66.4	6.5	1509	20	AAH25334	Feedback insensiti
35	66.4	6.5	1599	20	AAH211201	Mutant threonine d
36	66.4	6.5	1599	20	AAH25336	Feedback insensiti
37	66.4	6.5	1620	20	AAH211200	Mutant threonine d
38	66.4	6.5	1620	20	AAH25335	Feedback insensiti
39	66.4	6.5	1638	20	AAH211205	Mutant threonine d
40	66.4	6.5	1638	20	AAH25340	Feedback insensiti
41	66.4	6.5	1779	20	AAH211197	Mutant threonine d
42	66.4	6.5	1779	20	AAH211209	Wild type threonin
43	66.4	6.5	1779	20	AAH25331	Arabidopsis wild-t
44	66.4	6.5	1779	20	AAH25332	Feedback insensiti
45	66.4	6.5	1830	20	AAH211198	Mutant threonine d

ALIGNMENTS

RESULT 1  
ID AAA59294 standard; DNA; 1018 BP.

XX AAA59294;

XX 07-NOV-2000 (first entry)

XX DNA encoding a murine serine racemase polypeptide.

DE Serine racemase; N-methyl-D-aspartate receptor; neural death;

KW neural dysfunction; NMDA receptor; Parkinson's disease;

KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.

OS Mus musculus.

PH Key Location/Qualifiers

FT CDS 1..1018

FT /\*tag= a

FT /transl\_except= (pos: 1012, aa: Ser)

FT /product= "serine racemase"

XX WO200043526-A1.

XX PD 27-JUL-2000.

XX PF 18-JAN-2000; 2000WO-US00938.

XX PR 19-JAN-1999; 99US-0116333.

XX PR 21-JUL-1999; 99US-0144839.

XX PR 28-JUL-1999; 99US-0145953.



XX The present sequence encodes a mammalian serine racemase, which has  
CC a specific activity of at least 0.003 micromole L-serine/mg/hour.  
CC The enzyme catalyses the direct racemisation of L-serine to D-serine.  
CC D-serine appears to be an endogenous ligand of N-methyl-D-aspartate  
CC (NMDA) receptors. The mammalian serine racemases can be used to identify  
CC modulators, which can be used in the treatment of acute or chronic  
CC neural death or dysfunction mediated by overactivation of N-methyl-D-  
CC aspartate (NMDA) receptors. Overactivation of the receptors is  
CC associated with Parkinson's disease, Huntington's disease, motor neurone  
CC disease and Alzheimer's disease.  
XX  
XX Sequence 1672 BP; 491 A; 394 C; 379 G; 408 T; 0 other;  
SQ

Query Match 99.3%; Score 1011.2; DB 21; Length 1672;  
Best Local Similarity 99.7%; Pred. No. 5.9e-312;  
Matches 1013; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTGTCAGTACTGATCTCTTCTGCTGATGTTGAAAAAGCTCATATCAACATTCAA 60  
DB 219 ATGTGTCAGTACTGATCTCTTCTGCTGATGTTGAAAAAGCTCATATCAACATTCAA 278  
QY 61 GACTCTATCCACCTCACCCAGTCTAACAAGCTCATTTTGAATCAATAGCAGGCGC 120  
DB 279 GACTCTATCCACCTCACCCAGTCTAACAAGCTCATTTTGAATCAATAGCAGGCGC 338  
QY 121 AATCTTTTCTTCAAAATGTAGCTCTTCCAGAAAACCTGGGTCTTTTAAAGATTCGAGGTGCC 180  
DB 339 AATCTTTTCTTCAAAATGTAGCTCTTCCAGAAAACCTGGGTCTTTTAAAGATTCGAGGTGCC 398  
QY 181 CTTAATGCCATCAGAGGCTTAATTCCTGACAGCCAGAGAGAGCCCAAGCCCGTAGTT 240  
DB 399 CTTAATGCCATCAGAGGCTTAATTCCTGACAGCCAGAGAGAGCCCAAGCCCGTAGTT 458  
QY 241 ACTCACAGCGGGAACCATGGCCAAAGCTCTCACTATGCTGCTAAATCGAAGGAAT 300  
DB 459 ACTCACAGCGGGAACCATGGCCAAAGCTCTCACTATGCTGCTAAATCGAAGGAAT 518  
QY 301 CTGCTTACATGTGGTTTCCCAAAACAGCTCCCACTCCAGAAACTGGCAATCCAAGCC 360  
DB 519 CTGCTTACATGTGGTTTCCCAAAACAGCTCCCACTCCAGAAACTGGCAATCCAAGCC 578  
QY 361 TATGAGCATCGATAGTATGCTGACCAAGTGCAGAGTGCAGAGAAAGGTCACTCAA 420  
DB 579 TATGAGCATCGATAGTATGCTGACCAAGTGCAGAGTGCAGAGAAAGGTCACTCAA 638  
QY 421 AGAATTATGCAAGAAACAGAGGATCTTGGTCCATCCCAACAGAGGCTGCAGTGATA 480  
DB 639 AGAATTATGCAAGAAACAGAGGATCTTGGTCCATCCCAACAGAGGCTGCAGTGATA 698  
QY 481 GCTGACAAGGAACAATTCCTGGAAGTGTGTAACCAAGGTTCCCTTGGTAGATCACTG 540  
DB 699 GCTGACAAGGAACAATTCCTGGAAGTGTGTAACCAAGGTTCCCTTGGTAGATCACTG 758  
QY 541 GTGGTACAGTAGGAGGAGGAGGATGTTGCTGGAATAGCCATTACAAATTAAGCCCTG 600  
DB 759 GTGGTACAGTAGGAGGAGGAGGATGTTGCTGGAATAGCCATTACAAATTAAGCCCTG 818  
QY 601 AAACCTAGTGTGAAGTATAGCTGCTGAGCCCTCGAATGAGATGAGTGTACCAAGTCT 660  
DB 819 AAACCTAGTGTGAAGTATAGCTGCTGAGCCCTCGAATGAGATGAGTGTACCAAGTCT 878  
QY 661 AAATGAAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGGTGC 720  
DB 879 AAATGAAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGGTGC 938  
QY 721 AAATCCAGATGGCTTGAATACCTGGCTTATATAAGAGACCTTGTGGATGATGCTTC 780  
DB 939 AAATCCAGATGGCTTGAATACCTGGCTTATATAAGAGACCTTGTGGATGATGCTTC 998  
QY 781 ACTGTCACCGAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGGGAGAACTG 840  
DB 999 ACTGTCACCGAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGGGAGAACTG 1058

QY 841 CTCATTGAGCCGACTGCTGGCTGGCAGTGGCTGCTCAGCTGCTCTCAGCATTTCCAAACA 900  
DB 1059 CTCATTGAGCCGACTGCTGGCTGGCAGTGGCTGCTCAGCTGCTCTCAGCATTTCCAAACA 1118  
QY 901 GTCTCTCCAGAAAGTAAGAAGCTCTGCAATGTACTAGTGGGGGAATGTAGACCTAACCC 960  
DB 1119 GTCTCTCCAGAAAGTAAGAAGCTCTGCAATGTACTAGTGGGGGAATGTAGACCTAACCC 1178  
QY 961 TCCCTGAACTGGGTGGGCGAGCTCAACGGCCAGCTCTTACCAGACGCTCTGTTT 1016  
DB 1179 TCCCTGAACTGGGTGGGCGAGCTCAACGGCCAGCTCTTACCAGACGCTCTGTTT 1234

RESULT 3  
AAA59299  
ID AAA59299 standard; DNA; 1023 BP.  
XX  
AC AAA59299;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE DNA encoding a human serine racemase polypeptide.  
XX  
KW Serine racemase; N-methyl-D-aspartate receptor; neural death;  
KW Huntington's disease; NMDA receptor; Parkinson's disease;  
KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1023  
FT /\*tag= a  
FT /product= "serine racemase"  
XX  
PN WO200043526-A1.  
PD 27-JUL-2000.  
XX  
PF 18-JAN-2000; 2000WO-US00938.  
XX  
PR 19-JAN-1999; 99US-0116333.  
PR 21-JUL-1999; 99US-0144839.  
PR 28-JUL-1999; 99US-0145953.  
XX  
PA (UWJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
XX  
PI Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;  
PI Ferris CD;  
XX  
DR WPI; 2000-482915/42.  
DR P-PSDB; AAB07734.  
XX  
PT Mammalian serine racemase preparations, used to identify modulators  
PT which can be used to treat diseases associated with  
PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease  
XX  
PS Claim 19; Page 49-50; 54pp; English.  
XX  
CC The present sequence encodes a mammalian serine racemase, which has  
CC a specific activity of at least 0.003 micromole L-serine/mg/hour.  
CC The enzyme catalyses the direct racemisation of L-serine to D-serine.  
CC D-serine appears to be an endogenous ligand of N-methyl-D-aspartate  
CC (NMDA) receptors. The mammalian serine racemases can be used to identify  
CC modulators, which can be used in the treatment of acute or chronic  
CC neural death or dysfunction mediated by overactivation of N-methyl-D-  
CC aspartate (NMDA) receptors. Overactivation of the receptors is  
CC associated with Parkinson's disease, Huntington's disease, motor neurone  
CC disease and Alzheimer's disease.  
XX  
SQ Sequence 1023 BP; 290 A; 220 C; 243 G; 270 T; 0 other;

Query Match		79.4%; Score 807.8; DB 21; Length 1023;
Best Local Similarity		87.7%; Pred. No. 4.1e-247;
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;		
QY	1	ATGTGTGTCAGTACTGTCATCTCCTTTTGGTGTGATGTTGAAAAAGCTCATATCAACATCAA 60
DB	1	ATGTGTGTCAGTACTGTCATCTCCTTTTGGTGTGATGTTGAAAAAGCTCATATCAACATCGA 60
QY	61	GACTCTATCCACTCACCCAGTGTCTAACAAGCTCCATTTGAATCAATAGCAGGGGC 120
DB	61	GACTCTATCCACTCACCCAGTGTCTAACAAGCTCCATTTGAATCAATAGCAGGGGC 120
QY	121	AATCTTTTCTTCAATGTGAGCTCTTCCAGAAACTGGCTCTTTAAAGATTCGAGGTGCC 180
DB	121	AATCTTTTCTTCAATGTGAGCTCTTCCAGAAACTGGCTCTTTAAAGATTCGAGGTGCC 180
QY	181	CTTAATGCCATCAGAGGCTTAATCTCTGACAGCCGAGAGAGAGCCCAAGCCGTAGTT 240
DB	181	CTTAATGCCATCAGAGGCTTAATCTCTGACAGCCGAGAGAGAGCCCAAGCCGTAGTT 240
QY	241	ACTCAGAGCGGGAAGCAATGCGCAAGCTCTCAGCTATGCTGCTAACTGGAAGGAAT 300
DB	241	ACTCAGAGCGGGAAGCAATGCGCAAGCTCTCAGCTATGCTGCTAACTGGAAGGAAT 300
QY	301	CCTGCTTACATGTGTGTTCCCAAAAGCTCCCAACTGCAAGAACTGGCAATCCAAGCC 360
DB	301	CCTGCTTATATGTGTGTTCCCAAGCTCCCAACTGCAAGAACTGGCAATCCAAGCC 360
QY	361	TATGAGCATCATAGTATCTGTGACCCAGTACGAGTCCAGAGAGAAAGGTCACTCAA 420
DB	361	TACGGAGGCTCAATTTGTACTGTGAACCTAGTATGATGAGTCCAGAGAGAAATGTTCAAAA 420
QY	421	AGAATTTATCAGAAACAGAGGCAATCTGTCATCCATCCCAAGGAGCTGCACTGATA 480
DB	421	AGATTTATCAGAAACAGAGGCAATCTGTCATCCATCCCAAGGAGCTGCACTGATA 480
QY	481	GCTGACAGGAACAATGCTGCTGGAAGTGTGTAACCAAGTTCCTTGGTATGATCACTG 540
DB	481	GCTGACAGGAACAATGCTGCTGGAAGTGTGTAACCAAGTTCCTTGGTATGATCACTG 540
QY	541	GTGGTACCATAGGAGGAGGAGGAATGTTGCTGGAATAGCCATACAAATTAAGGCCCTG 600
DB	541	GTGGTACCTGTAGTGGAGGAGGAATGTTGCTGGAATAGCAATTAAGGCTCTG 600
QY	601	AAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTGCAATGCAAGTACTGCTACCACT 660
DB	601	AAACCTAGTGTGAAGGTATATGCTGCTGCAACCTCAATGCAAGTACTGCTACCACT 660
QY	661	AAACTGAAAGGAGAACTGATGCCCAATCTTTATCTCCAGAAACCATGAGATGGTCTC 720
DB	661	AAAGCTGAAAGGAGAACTGATGCCCAATCTTTATCTCCAGAAACCATGAGATGGTCTC 720
QY	721	AAATCCAGATGGCTTGAATACCTGGCTTATTAAGAGACCTTGTGGATGATGCTTTC 780
DB	721	AAATCCAGATGGCTTGAATACCTGGCTTATTAAGAGACCTTGTGGATGATGCTTTC 780
QY	781	ACTGTCCAGGAGTGAATCAAGTATGCAACCCAGCTGTTGGGGGAGAGTGAACCTG 840
DB	781	ACTGTCCAGGAGTGAATTAAGTGTGCAACCCAGCTGTTGGGGGAGAGTGAACCTG 840
QY	841	CTCATTGAGCCGACTGCTGGGTGGCACTGGCTGAGTGTCTCTCAGCATTTTCAACA 900
DB	841	CTCATTGAGCCGACTGCTGGGTGGCACTGGCTGAGTGTCTCTCAGCATTTTCAACA 900
QY	901	GTCCTCCAGAGTAAAGCAAGCTGTGATTTGCTACGTGGGGGAAATGTAGACCTAA 958
DB	901	GTTTCCCAAGTAAAGCAATTTGTTATGTGCTGAGTGGGAAATGTAGACCTTACC 960
QY	959	-CCTCCCTGAACTGGGTGGGGAGGCTGAAGCGCCAGCTCCTTACCAGAGGCTGTGTT 1016
DB	961	TCTTCCATAACTTGGGTGAAGCAGGCTGAAGGCGCCAGCTTCTTATCATGCTGTGTTCTGT 1019

RESULT 4		
AAS15217		
ID	AAS15217	standard; cDNA; 1023 BP.
XX	AAS15217;	
AC	16-JAN-2002	(first entry)
DT	Human cDNA encoding Serine Racemase.	
DE	Human; ss; serine racemase; Parkinson's disease; Huntington's disease; anxiety; glaucoma; stroke; hyperalgesia; pain; spinocerebellar ataxia; schizophrenia; transgenic animal; chromosome 17p13.	
KW	Homo sapiens.	
OS	Key	Location/Qualifiers
XX	1..1023	
FT	/*tag= a	
FT	/product= "Serine racemase"	
XX	WO200175144-A1.	
PN	11-OCT-2001.	
XX	02-APR-2001; 2001WO-US10662.	
PF	04-APR-2000; 2000US-194451P.	
XX	(MERI ) MERCK & CO INC.	
XX	Connolly T, Liu Y, Xia M;	
PI	WPI: 2001-656991/75.	
XX	P-PSDB; AAU09124.	
DR	New recombinant serine racemase polypeptide, useful in assays for identifying compounds that alter enzyme activity (e.g. including compounds that inhibit or stimulate enzyme activity) or in generating antibodies against the protein	
PT	Claim 1; Page 9; 43pp: English.	
XX	The invention relates to a recombinant polynucleotide encoding a human serine racemase, vectors containing it, host cells expressing the racemase, methods of identifying inhibitors of serine racemase and a transgenic animal lacking a functional endogenous serine racemase comprising the human serine racemase of the invention. The racemase protein is useful in assays to identify compounds that inhibit or stimulate enzyme activity, in the generation of antibodies against the protein, and in structural studies of the protein and	
CC	structure/function relationships of the protein. Biologically active fragments, and mutant or polymorphic forms of the serine racemase polypeptide have diagnostic, therapeutic or prophylactic uses (e.g. for neurological diseases such as Parkinson's and Huntington's disease, anxiety, glaucoma, stroke, hyperalgesia, pain, spinocerebellar ataxia and schizophrenia), and would be useful for screening for modulators and/or inhibitors of serine racemase function. The polynucleotides are useful as probes for the specific detection of the presence of a polynucleotide encoding a serine racemase protein, and as primers for nucleic acid amplification based assays for the detection of polynucleotides encoding serine racemase protein. The transgenic animal is useful for the study of the tissue and temporal specific expression or activity of the serine racemase gene in an animal. The gene for human serine racemase is located on chromosome 17p13. The present sequence encodes human serine racemase.	
CC	Sequence 1023 BP; 290 A; 220 C; 243 G; 270 T; 0 other;	
SQ	Query Match 79.4%; Score 807.8; DB 22; Length 1023;	
Best Local Similarity 87.7%; Pred. No. 4.1e-247;		
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;		

```
QY 1 ATGTGTGCTCAGTACTGTCATCTCTCTTGTGTGATGTTGAAAAAGCTCATATCAACATTCAA 60
Db 1 ATGTGTGCTCAGTATTGTCATCTCTTGTGTGATGTTGAAAAAGCTCATATCAACATTCAA 60
QY 61 GACTCTATCCACCTCACCCAGTCTAACAAGCTCCATTTTGAATCAAAATAGCAGGGCC 120
Db 61 GATTCTATCCACCTCACACAGTGCTAACAAGTCCATTTTGAATCAAAATAGCAGGGCC 120
QY 121 AATCTTTTCTTCAAAATGTCAGCTCTTCCAGAAAACCTGGGCTTTTAAAGATTCCAGGTGCC 180
Db 121 AATCTTTTCTTCAAAATGTCAGCTCTTCCAGAAAACAGGATCTTTTAAAGATTCTCGTGTGCT 180
QY 181 CTTAATGCCATCAGAGGCTTAATCTCTGACACGCCAGAGAGAGCCAAAGCCGTAGTT 240
Db 181 CTTAATGCCGTCAGAGGCTTGGTCTCTGTATGCTTTAGAAAAGAGCGCAAGCTGTGTT 240
QY 241 ACTCAGCAGCGGAAACCATGCGCAAGCTCTCACCTATGCTCTAACTGGAAGGAATT 300
Db 241 ACTCAGCAGCTGGAACCATGCGCAGGCTCTCACCTATGCTCTGCAAAATGGAAGGAATT 300
QY 301 CTTGCTTACATTTGTTTCCCAACAGCTCTCCAACTGCAAGAACTGGCAATCCCAAGCC 360
Db 301 CTTGCTTATATTGTTGCGCCAGACAGCTCCAGACTGTAAAACTTGAATACAGCC 360
QY 361 TATGGAGATCGATAGTATCTGTCAGCCCAAGTGTGAGAGTCCAGAGAAAAGGTCACTCAA 420
Db 361 TACGGAGCGTCAATTTGATCTGTGAACCTAGTGTGATGATGATGATGATGATGATGATGAT 420
QY 421 AGAATTATGCAACAAACAGAGGCTCTGCTGCTCCATCCCAACAGGAGCGCTGAGTGATA 480
Db 421 AGATTACAGAAACAAAGAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATA 480
QY 481 GCTGGACAAGGAAATTTGCCCTGGAAGTGTCTGAACCAAGTTCCTTTGGTGTGATGACTG 540
Db 481 GCTGGACAAGGAAATTTGCCCTGGAAGTGTCTGAACCAAGTTCCTTTGGTGTGATGACTG 540
QY 541 GTGGTACCATGAGGAGAGGAAATGTTGCTGGAAATAGCCATTAATTAAGGCCCTG 600
Db 541 GTGGTACCTGTAGTGTGAGGAGGAAATGCTTGTGGAATAGCAATTAATTAAGGCTCTG 600
QY 601 AAACCTAGTGTGAGGTATATGCTGTGAGCCCTCGAATGCAATGCAATGCAATGCAATGCT 660
Db 601 AAACCTAGTGTGAGGTATATGCTGTGAGCCCTCGAATGCAATGCAATGCAATGCAATGCT 660
QY 661 AAATGAAAGGAGAACTGACCCCAATCTTCTATCTCCAGAAACCATAGCAGATGGTGTG 720
Db 661 AAGCTGAAAGGAGAACTGATGCCCAATCTTATCTCCAGAAACCATAGCAGATGGTGTG 720
QY 721 AAATCAGCATTTGGCTTGAATACCTGGCTTATTAAGAGACCTTGTGGATGATGCTTC 780
Db 721 AAATCAGCATTTGGCTTGAACACCTGGCTTATTAAGAGACCTTGTGGATGATGATGCTTC 780
QY 781 ACTGTCCAGCAAGATGAATCAAGTATGCAACCCAGCTGTGTGGGGAGAAATGAACATG 840
Db 781 ACTGTCCAGCAAGATGAATTAAGTGTGCAACCCAGCTGTGTGGGGAGAGATGAACATG 840
QY 841 CTCATTGAGCCGACTGTGGCTGGCACTGGCTGTCAGTGTCTGCTCAGCATTTTCCAAACA 900
Db 841 CTCATTGAACTTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAACT 900
QY 901 GTCTCTCCAGAACTAAGAACGCTGTGATGTACTCAGTGGGGGGAATGTAGACCTAA-- 958
Db 901 GTTTCCTCCAGAACTAAGAACATTTTGTATGTGCTCAGTGGTGGAAATGTAGACTTAAAC 960
QY 959 -CTCTCCTGAACCTGGGTGGGAGGCTGAACGCCAGCTCTTACCAGAGGCTCTTTT 1016
Db 959 TCTCTCAACTTGGGTGAAGCAGGCTGAAAGGCCAGCTCTTATCAGTCTGTTTCTGT 1019
```

RESULT 5  
AAD14461  
ID AAD14461 standard; cDNA; 1770 BP.  
XX

```
AC AAD14461;
XX 15-NOV-2001 (first entry)
XX Human pyridoxal-phosphate dependent enzyme 22406 cDNA.
DE Human; pyridoxal phosphate dependent enzyme; nootropic; neuroprotective;
XX anticonvulsant; cerebroprotective; cardiant; vasotropic; gene therapy;
KW epilepsy; Alzheimer's disease; chromosome 17; serine racemase; stroke;
KW behavioural change; neurodegenerative disorder; schizoprenia; atresia;
KW rheumatic heart failure; circulatory disorder; hepatic injury; jaundice;
KW lung disorder; nodular hyperplasia; stenosis; skeletal muscle disorder;
KW tumour; rhabdomyosarcoma; dermal fibroblast disorder; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FT CDS 69..1091
FT /tag= a
FT /product= "Human pyridoxal phosphate dependent enzyme"
FT misc_feature 69..1088
FT /tag= b
FT /note= "This region is specifically claimed as
FT SED ID NO: 3 in claim 1"
XX WO200160987-A1.
XX 23-AUG-2001.
XX 20-FEB-2001; 2001WO-US05365.
XX 17-FEB-2000; 2000US-0183208.
XX (MILL-) MILLENNIUM PHARM INC.
XX Meyers RA, Rudolph-Owen LA;
XX WPI; 2001-529909/58.
XX P-PSDB; AAE08342.
XX Novel polypeptide of the human pyridoxal phosphate dependent family
XX useful in screening and detection assays and for treatment, e.g. of
XX epilepsy and Alzheimer's
XX Claim 1; Fig 1; 12lpp; English.
XX The present sequence is a cDNA encoding human pyridoxal phosphate
XX dependent enzyme 22406 which is a serine racemase. Human 22406 gene is
XX located on chromosome 17 between D17S849 and D17S796. The protein 22406
XX is a modulator of D-serine. D serine has been shown to modify
XX behavioural changes associated with learning, memory and convulsions.
XX Human 22046 and compounds that modulate the expression or activity are
XX used to treat or diagnose neurodegenerative disorders including
XX Alzheimer's disease, schizophrenia as well as quell anxiety and
XX epilepsy and prevent damage from stroke as well as cardiac (heart
XX failure, rheumatic heart failure) and circulatory disorders, liver
XX disorders (hepatic injury, jaundice), lung disorders, prostate
XX disorders (benign enlargement, nodular hyperplasia), colon disorders
XX (atresia, stenosis), skeletal muscle disorders (tumours-rhabdomyosarcoma)
XX and dermal fibroblast disorders. Human 22406 cDNA is also useful in gene
XX therapy.
XX Sequence 1770 BP; 499 A; 379 C; 410 G; 482 T; 0 other;
XX
XX Query Match 79.4%; Score 807.8; DB 22; Length 1770;
XX Best Local Similarity 87.7%; Pred. No. 5.5e-247;
XX Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;
QY 1 ATGTGTGCTCAGTACTGTCATCTCTTGTGTGATGTTGAAAAAGCTCATATCAACATTCAA 60
Db 69 ATGTGTGCTCAGTATTGTCATCTCTTGTGTGATGTTGAAAAAGCTCATATCAACATTCAA 128
QY 61 GACTCTATCCACCTCACCCAGTCTAACAAGCTCCATTTTGAATCAAAATAGCAGGGCGC 120
```

Db 129 GATTCTATCCACCTCACACAGTGTCTAACAGTCCATTTTGAATCAACTAACAGGCGC 188  
Qy 121 AATCTTTTCTCAAACTGTAGCTCTTCCAGAAACTGGTCTTTTAAGATTCGAGTGCC 180  
Db 189 AATCTTTTCTCAAACTGTAGCTCTTCCAGAAACTGGTCTTTTAAGATTCGAGTGCC 248  
Qy 181 CTTAATGCCATCAGAGCTTTAATCTCTGACAGCCGAGAGAGAGCCCAAGCCGTAGTT 240  
Db 249 CTCATGCCCTCAGAACCTTGGTTCCTGTGATGCTTTAGAAAGGAAGCCGAAGCTGTGTT 308  
Qy 241 ACTCAGCAGGAGGAACCATGCGCAAGCTCTCACCTATGCTGCTAACTGGAGGAATT 300  
Db 309 ACTCAGCAGGAGGAACCATGCGCAAGCTCTCACCTATGCTGCTAACTGGAGGAATT 368  
Qy 301 CTTGCTTACATGTGTTTCCCAACAGCTCCCAACTGCAAGAACTGGCAATCCCAAGCC 360  
Db 369 CTTGCTTATATGTGTTGCGCCAGACAGCTCCAGACTGTAAACAACTGCAATACAGCC 428  
Qy 361 TATGGAGCATCGATAGTATATCTGTGACCCAAAGTGACAGTCCAGAGAAAAGTCACTCAA 420  
Db 429 TACGGAGCGTCAATTTGATGTAACCTAGTGTGATGAGTCCAGAGAAAATGTTGCAAAA 488  
Qy 421 AGAATTATGCAAGAAAGGATCTTGGTCCATCCCAACAGGAGCCCTGCAGTGATA 480  
Db 489 AGAGTTTACAGAAAGAAACAGAGGATCATGTGTATATCCCAACAGGAGCCCTGCAGTGATA 548  
Qy 481 GCTGGCAAGGAAACAATTTGCCCTGGAGTGTCTGAACCCAGCTTCCCTGGTGTAGTGCACCTG 540  
Db 549 CTTGGACAGGGACAATTTGCCCTGGAGTGTCTGAACCCAGCTTCCCTGGTGTAGTGCACCTG 608  
Qy 541 GTGTTACAGTAGGAGGAGGAATGTTGCTGGAATAGCCATTACAATTAAGGCCCTG 600  
Db 609 GTGTTACCTCTAGTGTGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTCTG 668  
Qy 601 AATCCTAGTGTAGGATATAGTGTGAGCCCTCGAATCGAGATGACTGCTACCAAGTCT 660  
Db 669 AATCCTAGTGTAGGATATATGCTGTGTAACCCCTCAATGCAGATGACTGCTACCAAGTCT 728  
Qy 661 AATCTGAAGGAGAACTGACCCCAATCTTTCATCTCCAGAAACCATAGCAGATGTTGTC 720  
Db 729 AGCTGAAGGAGAACTGATGCCCACATCTTATCTCCAGAAACCATAGCAGATGTTGTC 788  
Qy 721 AATCAGCATTTGGCTTGAATACCTGGCCCTATTATTAAGAGACCTTGTGGATGATGCTTC 780  
Db 789 AATCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGACCTTGTGGATGATGATCTTC 848  
Qy 781 ACTGTCAACCAAGATGAATCAAGTATGCAACCCAGCTGTTGTGGGAGAAATGAACCTG 840  
Db 849 ACTGTCAACCAAGATGAATTAAGTGTGCAACCCAGCTGTTGTGGGAGAGGATGAACCTA 908  
Qy 841 CTCATTGAGCCGACTGCTGGCGTGGCACTGGCTGCAGTGTGCTCAGCATTTTCCAAACA 900  
Db 909 CTCATTGAACCTACAGCTGTTGTTGGAGTGGCTGCTGTCTCTCAACATTTTCAACT 968  
Qy 901 GTCTCTCCAGAAGTAAGAACGCTGCTGATTTGTAAGTGGGGGGAATGTAGACCTAA-- 958  
Db 969 GTTTCGCCAGAGTAAGAACATTTGTTATGCTCAGTGGTGAATGTAGACTTAACC 1028  
Qy 959 -CTCCCTGNACTGGGTGGGCGAGCTGACGCGCCAGCTCCTTACCAGACGGTCTGTT 1016  
Db 1029 TCCTCCATAACTTGGGTGAAGCAGGCTGAAGGCCAGCTTCTTATCAGTCTGTTTCTGT 1087

RESULT 6  
AAH16282  
ID AAH16282 standard; cDNA; 2477 BP.  
XX  
AC AAH16282;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:15148.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
KW Homo sapiens.  
XX  
XX EPI074617-A2.  
XX  
XX 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
XX 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Isodai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
DR  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 15148; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 2477 BP; 738 A; 535 C; 524 G; 680 T; 0 other;

Query Match 79.4%; Score 807.8; DB 22; Length 2477;  
Best Local Similarity 87.7%; Pred. No. 6.6e-247;  
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;  
Qy 1 ATGTGTGCTCAGTACATCTCTTTGCTGTGATGTTTGAAGAAAGCTCATATCAACATTCAA 60  
Db 69 ATGTGTGCTCAGTATTCATCTCTTTGCTGTGATGTTTGAAGAAAGCTCATATCAACATTCAA 128  
Qy 61 GACTCTATCCACCTCACCCAGTGCCTAACAGCTCCATTTTGAATCAATAGCAGGCGC 120  
Db 129 GATTCTATCCACCTCACCCAGTGCCTAACAGCTCCATTTTGAATCAACCTAACAGGCGC 188  
Qy 121 AATCTTTTCTCAAACTGTAGCTCTTCCAGAAACTGGTCTTTTAAGATTCGAGTGCC 180  
Db 189 AATCTTTTCTCAAACTGTAGCTCTTCCAGAAACTGGTCTTTTAAGATTCGAGTGCC 248  
Qy 181 CTTAATGCCATCAGAGCTTTAATCTCTGACAGCCGAGAGAGCCCAAGCCGTAGTT 240

Db 249 CTCAATCGCGTCAGAGCTGGTCTCTGATGCTTTAGAAAGGAGCCGAAAGCTGTGTTT 308  
Qy 241 ACTCACAGCAGGAAACCATGGCCAAAGCTCTCACCTATGCTGCTTAAACTGGAAGGAATT 300  
Db 309 ACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGAAGGAATT 368  
Qy 301 CCTGCTTACATGTGGTGGTCCCAACACAGCTCCCAACTGCAAGAACTGGCAATCCAGCC 360  
Db 369 CCTGCTTATATGTGGTCCCAACAGAGCTCCAGACTGTAAAGAACTTTGCAATACAGCC 428  
Qy 361 TATGAGCATCGATAGTATATGTCAGCCCAAGTGCAGAGTCCAGAGAAAGGTCACTCAA 420  
Db 429 TACGAGCGCTCAATTTGATCTGTGAACCTAGTATGATGATGATGATGATGATGATGATG 488  
Qy 421 AGAATTTATGCAAGAAACAGAGGATCTTGTGTCATCCCAACAGGAGCCCTGCAGTGATA 480  
Db 489 AGAGTTACAGAAAGAAACAGAGGATCATGTACATCCCAACAGGAGCCCTGCAGTGATA 548  
Qy 481 GCTGACAAAGAAACAAATTTGCCCTGGAAGTGTGTAACAGGTTCCCTTGGTATGATGCTG 540  
Db 549 GCTGACAAAGGACAATTTGCCCTGGAAGTGTGTAACAGGTTCCCTTGGTATGATGCTG 608  
Qy 541 GTGGTACCACTAGGAGGAGGAGGATGTTGCTGGAATAGCCATTACAAATTAAGCCCTG 600  
Db 609 GTGGTACCTGTAGTGGAGGAGGAGGATGTTGCTGGAATAGCAATTTACAGTTAAGGCTG 668  
Qy 601 AAACCTAGTGTGAAGTATACGCTCTGAGCCCTGGAATGAGATGAGTGTACCAAGTCT 660  
Db 669 AAACCTAGTGTGAAGTATATGCTGTAACCCCTCAATGAGATGAGTGTACCAAGTCT 728  
Qy 661 AAACCTGAAGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGGTCTC 720  
Db 729 AAGCTGAAGGGAACACTGATGCCAATCTTTATCTCTCCAGAAACCATAGCAGATGGTCTC 788  
Qy 721 AAATCCACATGCTGCTCAATACCTGGCTTATATATAGAGACCTTGTGGATGATGCTTC 780  
Db 789 AAATCCACATGCTGCTCAATACCTGGCTTATATATAGAGACCTTGTGGATGATGCTTC 848  
Qy 781 ACTGTCACCGAAGTGAATCAAGTATGCAACCCAGCTGGTGGGGAGAGATGAACCTG 840  
Db 849 ACTGTCACAGAGGATGAATTAAGTGTCAACCCAGCTGGTGGGGAGAGATGAACCTA 908  
Qy 841 CTCAATTAGCCGACGTGGGCTGGCTGCACTGCTGCAAGTGTGCTGCACTATTTCAACA 900  
Db 909 CTCAATTGAACCTACAGCTGGTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 968  
Qy 901 GTCCTCCAGAGTAAAGCACTGTCATGTCAGTGGGGGGAATGTAGACCTAA-- 958  
Db 969 GTTCCCCAGAGTAAAGCAATTTGTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1028  
Qy 959 -CCTCCCTGAACGTGGTGGGAGGCTGAACGGCCAGCTCTTACAGACAGCTGTGTTT 1016  
Db 1029 TCCCTCAATACTGGGTGAAGCAGCTGAAGGCCAGCTTCTTATCAGCTGTGTTCTGT 1087

RESULT 7

AAI70575  
ID AAI70575 standard; cDNA; 1336 BP.  
XX AC AAI70575;  
XX DT 21-JAN-2002 (first entry)  
XX DE Human serine racemase cDNA.  
XX KW Serine racemase; human; D-serine; regulation;  
KW glutamate N-methyl-D-aspartate receptor; neurodegenerative disease;  
KW stroke; Alzheimer's disease; Parkinson's disease; pain; myopathy;  
KW nootropic; neuroprotective; cerebroprotective; antiparkinsonian;  
KW analgesic; diagnosis; gene therapy; screening; ss.  
XX OS Homo sapiens.

Query Match 79.2%; Score 806.2; DB 22; Length 1336;  
Best Local Similarity 87.6%; Pred. No. 1.5e-246;  
Matches 893; Conservative 0; Mismatches 123; Indels 3; Gaps 1;

Qy 1 ATGCTGCTCAGTACTGTCATCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCAA 60  
Db 52 ATGCTGCTCAGTATTTGTCATCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCGA 111  
Qy 61 GACTCTATCCACCTCACCCAGTGTCTTCCAGAAAACTGGTCTTTTAAGATTCAGAGGCC 120  
Db 112 GATTCTATCCACCTCACACCAGTGTCTTCCAGAACTCCATTTTCAATCAACTAACAGGCC 171  
Qy 121 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAAACTGGTCTTTTAAGATTCAGAGGCC 180  
Db 172 AATCTTTTCTTCAATGTGAGCTCTTCCAGAACTCCATTTTCAAGATTCAGAGGCC 231  
Qy 181 CTTAATGCCATCAGAGGCTTAACTTCTGACACGCCAGAGAGAACCCCAAGCCCTAGT 240  
Db 232 CTCATGCCGTCAGAGGCTTGGTCTTCTGATGCTTTTAGAAGGAGGAGGAGGAGGCTGT 291

XX Key Location/Qualifiers  
FH CDS 52..1074  
ET /\*tag= a  
XX WO200173077-A2.  
XX 04-OCT-2001.  
XX 30-MAR-2001; 2001WO-EP03668.  
XX 31-MAR-2000; 2000US-193748P.  
XX 03-APR-2000; 2000US-194249P.  
XX (FARB ) BAYER AG.  
XX Ramakrishnan S;  
XX WPI; 2001-648444/74.  
XX P-PSDB; AAM50262.  
XX Polynucleotide encoding serine racemase enzyme and the enzyme useful  
XX for screening reagents regulating the activity of the enzyme in a  
XX neuron disease caused by over- or under-activation of glutamate  
XX N-methyl-D-aspartate  
XX Claim 1; Fig 1; 66pp; English.  
XX The present sequence is that of cDNA encoding human serine racemase  
XX (see AAM50262). The polynucleotide can be used in the production  
XX of recombinant serine racemase enzyme, or for the detection of  
XX serine racemase polynucleotides. Expression vectors and host cells  
XX are claimed. Serine racemase catalyses the conversion of L-serine  
XX to D-serine. Neuron damage following various nervous system diseases  
XX is often caused by activation of glutamate N-methyl-D-aspartate  
XX (NMDA) receptors in the brain. This receptor is activated by the  
XX binding of D-serine. Regulation of D-serine levels through  
XX regulation of serine racemase may therefore prevent or minimize  
XX neuron damage in neurogenic and myopathic disorders,  
XX neurodegenerative disorders such as Alzheimer's disease and  
XX Parkinson's disease, and disorders leading to peripheral and  
XX chronic pain. Serine racemase polypeptides and polynucleotides are  
XX used in claimed methods of screening for agents that modulate or  
XX decrease the activity of serine racemase. Also claimed is a  
XX pharmaceutical composition comprising either an expression vector  
XX that contains a serine racemase polynucleotide, or a reagent that  
XX modulates serine racemase enzyme activity. This is used to modulate  
XX serine racemase activity in a disease, particularly neuron damage or  
XX a neurodegenerative disease caused by the over- or under-activation  
XX of the glutamate NMDA receptor.

QY 241 ACTCACAGCAGCGGAAACACATGCCCCAGCTCTCACCTATGCTGCTAACTGGAAGAAATT 300  
Db 292 ACTCACAGCAGTGAACACACATGCCCCAGCTCTCACCTATGCTGCTGCAAAATTGGAAGAAATT 351  
QY 301 CCTGCTTACATTTGTGTTTCCCAACAGAGCTCCCACTGCAAGAAATGCAATGCCAAGCC 360  
Db 352 CCTGCTTATATTTGTGTTGCCCAAGAGCTCCAGAGCTGTAAGAAATGCAATGCCAAGCC 411  
QY 361 TATGAGACATCATAGTATACCTGACCCAGAGTCCAGAGTCCAGAGAAAGTCACTCAA 420  
Db 412 TACGGAGCTCAATTTGTATACCTGAGAGTCCAGAGTCCAGAGAAATGTCAGAAA 471  
QY 421 AGAATTATGCAAGAAACAGAGCATCTTTGGTCCATCCCAACAGAGCTCCAGAGTGA 480  
Db 472 AGAGTTACAGAGAAACAGAGCATCTGTTATACATCCCAACAGAGCTCCAGAGTGA 531  
QY 481 GCTGGACAAGAACAAATTCCTGGAAGTCTGGAACAGAGTCCCTTTGGTAGAGTCACTG 540  
Db 532 GCTGGACAAGGACAAATTCCTGGAAGTCTGGAACAGAGTCCCTTTGGTAGAGTCACTG 591  
QY 541 GTGGTACCACTAGGAGGAGGAGAAATGTTGCTGGAATAGCAATACCAATTAAGCCCTG 600  
Db 592 GTGGTACCACTAGGAGGAGGAGAAATGTTGCTGGAATAGCAATACCAATTAAGCCCTG 651  
QY 601 AAACCTAGTGTGAAGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCTACCACTCT 660  
Db 652 AAACCTAGTGTGAAGTATATGCTGGAACCTCAAAATGCAGATGACTGCTACCACTCT 711  
QY 661 AAACCTAGGAGAACTGACCCCAATCTTCATCTCCAGAACCAATAGCAGATGCTGTC 720  
Db 712 AAGCTGAAGGGGAACTGATGCCCCAATCTTTATCTCCAGAACCAATAGCAGATGCTGTC 771  
QY 721 AAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATAAGAGACCTTGTGGATGATGCTTC 780  
Db 772 AAATCCAGCATTTGGCTTGAACACTGGCCCTATTATAAGAGACCTTGTGGATGATGCTTC 831  
QY 781 ACTGTACCGAAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGGAGAGTGAACACTG 840  
Db 832 ACTGTACAGAGATGAATTAAGTGTGCAACCCAGCTGGTGTGGGGAGAGTGAACACTG 891  
QY 841 CTCATTGAGCCGCTGCTGGCTGACCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
Db 892 CTCATTGAGCCGCTGCTGGCTGACCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 951  
QY 901 GTCTCTCCAGAGTAAAGAACTGCTCAATTTGCTGAGTGGGGGAAATGAGACCTAA-- 958  
Db 952 GTTTCCTCCAGAGTAAAGAACTTTGCTGCTGAGTGGGGGAAATGAGACCTAA-- 1011  
QY 959 -CCTCCCTGAAGTGGTGGGGGAGGCTGAAAGCCAGCTGCTTACAGAGCTGCTGTTT 1016  
Db 1012 TCCTCCATAAATTTGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTCTGT 1070

RESULT 8

AAI65020  
ID AAI65020 standard; cdNA; 2674 BP.

XX AC AAI65020;

DT 27-NOV-2001 (first entry)

DE Serine/threonine dehydrase 37 coding sequence.

XX Serine/threonine dehydrase 37; cancer; HIV infection; anti-HIV;  
KW cytosolic; ss.

OS Unidentified.

XX Key Location/Qualifiers  
FH 262..1284  
FT CDS

FT /tag= a  
FT /product= "Serine/threonine dehydrase 37"

PN CN1300824-A.  
XX 27-JUN-2001.  
XX 21-DEC-1999; 99CN-0125662.  
XX 21-DEC-1999; 99CN-0125662.  
XX (UYFU-) UNIV FUDAN.  
PA Mao Y, Xie Y;  
PI WPI; 2001-530468/59.  
XX P-PSDB; AAG78808.  
XX Polypeptide-serine/threonine dehydrase 37 and polynucleotide for coding  
PT this polypeptide -  
XX Claim 6; Page 25-26 (Disclosure); 34pp; Chinese.  
PS The present sequence is the coding sequence for serine/threonine  
CC dehydrase 37. The dehydrase and its coding sequence are useful for  
CC treating diseases e.g. cancer and HIV infection.  
XX Sequence 2674 BP; 803 A; 581 C; 563 G; 727 T; 0 other;  
SQ  
Query Match 79.2%; Score 806.2; DB 22; Length 2674;  
Best Local Similarity 87.6%; Pred. No. 2.2e-246;  
Matches 893; Conservative 0; Mismatches 123; Indels 3; Gaps 1;  
QY 1 ATGTGCTGCTCAGTACTGCTATCCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCAA 60  
Db 262 ATGTGCTGCTCAGTACTGCTATCCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCGA 321  
QY 61 GACTCTATCCACCTCACCCAGTGTCTAACAGCTCCATTTTGAATCAATAGCAGGCGC 120  
Db 322 GATTCTATCCACCTCACCCAGTGTCTAACAGCTCCATTTTGAATCAATAGCAGGCGC 381  
QY 121 AATCTTTTCTTCAAAATGTGAGCTCTTCCAGAAAGCTGGTCTTTTAAGATTCGAGGTGC 180  
Db 382 AATCTTTTCTTCAAAATGTGAGCTCTTCCAGAAAGCTGGTCTTTTAAGATTCGAGGTGC 441  
QY 181 CTTAATGCCATCAGAGGCTTAACTCTGACACGCCAGAGAGAGGCCCAAGCCGCTAGTT 240  
Db 442 CTCATCGCGTCAGAGCTTGGTCTCTGATGCTTTAGAAAGCAAGCCGAAAGCTGTTGT 501  
QY 241 ACTCACAGCAGCGGAAACACAGCTCCAGCTCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 300  
Db 502 ACTCACAGCAGCGGAAACACAGCTCCAGCTCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 561  
QY 301 CCTGCTTACATTTGTGTTTCCCAACAGCTCCCAACAGCTCCCAACAGCTCCCAACAGCTCC 360  
Db 562 CCTGCTTACATTTGTGTTTCCCAACAGCTCCCAACAGCTCCCAACAGCTCCCAACAGCTCC 621  
QY 361 TATGAGACATCATAGTATACCTGACCCAGTCCAGAGTCCAGAGTCCAGAGAAAGTCACTCAA 420  
Db 622 TACGAGCGCTCAATTTGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 681  
QY 421 AGAATTATGCAAGAAACAGAGCATCTTTGGTCCATCCCAACAGAGCTCCAGAGTGA 480  
Db 682 AGAGTTACAGAGAAACAGAGCATCTTTGGTAGATCCCAACAGAGCTCCAGAGTGA 741  
QY 481 GCTGGACAAGGAAACAAATTTGCCCTGGAAGTGTCTGGAACAGAGTCCCTTTGGTAGATGCTG 540  
Db 742 GCTGGACAAGGAAACAAATTTGCCCTGGAAGTGTCTGGAACAGAGTCCCTTTGGTAGATGCTG 801  
QY 541 GTGGTACCACTAGGAGGAGGAGAAATGTTGCTGGAATAGCAATTAAGCCCTG 600  
Db 802 GTGGTACCACTAGGAGGAGGAGAAATGTTGCTGGAATAGCAATTAAGCCCTG 861  
QY 601 AAACCTAGTGTGAAGTATACGCTGCTGAGCCCTGGAATAGCAATTAAGCCCTG 660  
Db 862 AAACCTAGTGTGAAGTATACGCTGCTGAGCCCTGGAATAGCAATTAAGCCCTG 921

QY 661 AACTGAAGGAGAACTGACCCCAATCTTATCTCCAGAAACCATAGCAGATGGTGC 720  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 922 AAGCTGAAGGGGAACTGATGCGCAATCTTATCTCCAGAAACCATAGCAGATGGTGC 981  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 721 AATTCAGCATTTGGCTTGAATACCTGGCCTATTATTAAGAGACCTTGTGGATGCTTTC 780  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 982 AATTCAGCATTTGGCTTGAACACCTGGCCTATTATCAGGACCTTGTGGATGATATCTTC 1041  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 781 ACTGTCCAGCAATGAATCAAGTATGCAACCCAGCTGGTGTGGGGAGAAATGAACAG 840  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 1042 ACTGTCCAGCAATGAATTAAGTGTGCACCCAGCTGGTGTGGAGAGGATGAACATA 1101  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 841 CTCATTGAGCCGACTGCTGGCGTGGCACTGGCTGCACTGCTGCTCAGCATTTCCAAACA 900  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 1102 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTCAACATTTCAAACT 1161  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 901 GTCTCTCCAGCAATGAAGACGCTGCTCATGTGACTCAGTGGGGGAGTGTAGACCTAA-- 958  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 1162 GTTCTCTCAGAATGAAGACATTTGTTATGTCTCAGTGGTGGAAATGTAGACTTAACC 1221  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 959 -CCTCCTGCAACTGGGTGGGGAGGCTGAACGGCCAGCTCCTTACCAGACGGTCTGTTT 1016  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 1222 TCCTCCATACTTGGGTGAAGCAGGCTGAAGGCCAGCTTCTTATCAGTCTGTTCTGT 1280  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 RESULT 9  
 AAH06600  
 ID AAH06600 standard; cDNA; 848 BP.  
 XX  
 AC AAH06600;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA clone (5'-primer) SEQ ID NO:3435.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 27-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 1; SEQ ID 3435; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH93893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 848 BP; 236 A; 187 C; 214 G; 207 T; 4 other;  
 Query Match 58.9%; Score 599.6; DB 22; Length 848;  
 Best Local Similarity 87.6%; Pred. No. 1.1e-180;  
 Matches 687; Conservative 0; Mismatches 93; Indels 4; Gaps 3;  
 QY 1 ATGTGTGCTCAGTACTGCATCTCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCAA 60  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 69 ATGTGTGCTCAGTATTGTCATCTCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCGA 128  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 61 GACTCTATCCACCTCACCCAGTCTAAACAGCTCCATTTTGAATCAAAATAGCAGGGCGC 120  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 129 GATTCTATCCACCTCACCCAGTCTAAACAGCTCCATTTTGAATCAAAATAGCAGGGCGC 188  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 121 AATCTTTTCTCAAAATGTGAGCTCTTCCAGAAACTGGGCTCTTTTAAAGATTCGAGGTGCC 180  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 189 AATCTTTTCTCAAAATGTGAACTCTTCCAGAAACACAGGATCTTTTAAAGATTCGAGGTGCT 248  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 181 CTTAATGCCATCAGAGGCTTAATTCCTGCACACGCCAGAGAGGCCAAGCCGTAGTT 240  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 249 CTCATATGCCCTCAGAGCTTTGGTTCCTGATGCTTTAGAAAGAGGCGCAAGCTGTGTT 308  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 241 ACTCAGACGACGCGAAACCAATGGCCAAAGCTCTCACCTATGCTGCTAACTTGGGAAGAAAT 300  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 309 ACTCAGACGAGTGGAAACCAATGGCCAGGCTCTCACCTATGCTGCTCAAAATTTGGGAAGAAAT 368  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 301 CTGCTTACATTTGTGTTTCCCAACAGCTCTCCCAACTGCAAGAACTTGGCAATCCCAAGCC 360  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 369 CTGCTTATATTTGTTGGTCCCGCCAGACAGCTCCAGACTGTAAAAAATCTGCAATCAAGCC 428  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 361 TATGGACCATCATAGTATCTGACCCCAAGTGGACGAGTCCAGAGAAAGGTCACCTCAA 420  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 429 TACGGAGCGTCAATTTGTATCTGTGAACCTAGTGTAGTCCAGAGAGAAATGTTGCAAAA 488  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 421 AGAATATGCAAGAAACAGAGGCTCTTGGTCCATCCCAACAGGAGGCTGCAGTGATA 480  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 489 AGATTTACAGAAACAGAGAGGCTCATGTATCTCCCAACAGGAGGCTGCAGTGATA 548  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 481 GCTGGACAGGAACAATTTGCCCTGGAAGTCTGTAACCAAGGTTCCCTTGGTAGATGCACTG 540  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 549 GCTGGACAGGGACAAATTTGCCCTGGAAGTCTGTAACCAAGGTTCCCTTGGTAGATGCACTG 608  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 541 GTGTTACCAGTAGGAGGAGGAATGGTTGCTGGAAATAGCCATTACAATTAAGGCCCTG 600  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 609 GTGGTACCTGTAGTGTGGAGAAAGTCTTGTGGAAATAGCAATTAAGGCTTGTG 668  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 601 AAACCTAGTGTGAAGGTATACGCTGTGAGCCCTCGAATCGCATGATGACTCTTACCAGTCT 660  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 669 AAACCTAGTGTGAAGGTATATGCTGTGAACCTCAATCGCATGACTG-TTCCAGTCC 727  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 661 AAACCTGAAGGAGAACTGACCCCAATCTTATCTCCAGAAACCATAGCAGATGGTGC 720  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 728 AAGCTGAAGGGGAAACTGATGCGCAATCTTTATCTCTTCAGAA--CCATACAGATGGTGC 785  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 721 AATTCAGCATTTGGCTTGAATACCTGGCCTATTATTAAGAGACCTTGTGGATGATGCTTC 780  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 786 AATTCCTA-CATTGGCTTGAACACCTGGGCTTATTATCANGGACCTTGGNGATGATATCTTT 844

Qy 781 ACTG 784  
 Db 845 ACTG 848

RESULT 10  
 ABL90122  
 ID ABL90122 standard; cDNA; 731 BP.  
 XX  
 AC ABL90122;  
 XX  
 DT 24-MAY-2002 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 684.  
 XX  
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;  
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200190304-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001WO-US16450.  
 XX  
 PR 19-MAY-2000; 2000US-205515P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Birse CE, Rosen CA;  
 XX  
 DR WFI; 2002-122018/16.  
 DR P-PSDB; ABB89713.  
 XX  
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive, and  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders -  
 XX  
 PS Claim 4; SEQ ID NO 684; 2081pp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (ABL89449-ABL90853) and proteins  
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 731 BP; 204 A; 158 C; 184 G; 178 T; 7 other;

Query Match 52.1%; Score 530.4; DB 24; Length 731;  
 Best Local Similarity 87.7%; Pred. No. 1.3e-158;  
 Matches 598; Conservative 1; Mismatches 81; Indels 2; Gaps 2;

Qy 1 ATGTGTGCTCAGTACTGCTCCTTGTGCTGATGTTGAAAAAGCTCATATCAACATTCGA 60  
 |||||

Db 47 ATGTGTGCTCAGTACTGCTCCTTGTGCTGATGTTGAAAAAGCTCATATCAACATTCGA 105  
 Qy 61 GACTCTATCCACCTCACCCAGTGTCAACAGCTCCATTTTGAATCAATAGCAGGCGC 120  
 |||||  
 Db 107 GATTCTATCCACCTCACCCAGTGTCAACAGCTCCATTTTGAATCAATAGCAGGCGC 166  
 Qy 121 AATCTTTTCTTCAAAATGTGAGCTCTTCCAGAAACTGGGTCTTTTAAGATTCGAGGTGCC 180  
 |||||  
 Db 167 AATCTTTTCTTCAAAATGTGAGCTCTTCCAGAAACTGGGTCTTTAAGATTCGAGGTGCC 226  
 Qy 181 CTTAATGCCATCAGAGCTTTAATCTTGACACGCCAGAGAGGCCAAAGCCGTAGTT 240  
 |||||  
 Db 227 CTCATGCCCTCAGAAGCTTGGTTCCTGATGCTTTAGAAAAGGAGCGAAGCTGTGTT 286  
 Qy 241 ACTCAGCAGCAGGAACCATGCGCAAGCTCTCACCTATGCTCTAACTGGAAGGAATT 300  
 |||||  
 Db 287 ACTCAGCAGCTGGAACCATGCGCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAATT 346  
 Qy 301 CCTGCTTACATTTGTTGTTCCCAAGAGCTTCCCAACTGCAAGAACTGGCAATCCCAAGCC 360  
 |||||  
 Db 347 CCTGCTTATATTCTGTTGTTGCCCAAGAGCTTCCAGACTGTAAAAAACTTGCATACAGCC 406  
 Qy 361 TATGGAGCTCGATAGTATATCTGTGACCCCAAGTGTGAGAGTCCAGAGAAAAGTCACTCAA 420  
 |||||  
 Db 407 TACGGAGCGTCAATTTGTATATCTGTGAACCTAGTGTGATGAGTCCAGAGAAAATGTTGCAAAA 466  
 Qy 421 AGAATTATGCAAGAAACAGAGGATCTTGGTCCATCCCAACAGGAGCCCTGCAGTCAATA 480  
 |||||  
 Db 467 AGAGTTACAGAGAAAGAAAGGAGGATCATGGTACATCCCAACAGGAGCCCTGCAGTGAATA 526  
 Qy 481 GCTGGACAGGAAACAATTTGCCCTGGAAGTGTCTGAACAGGTTCCCTTGTGATGCACTG 540  
 |||||  
 Db 527 GCTGGACAGGAGCAATTTGCCCTGGAAGTGTCTGAACAGGTTCCCTTGTGATGCACTG 586  
 Qy 541 GTGGTACCAGTAGCAGGAGGAGGAATGTTGCTGGAATAGCCATTAACAATTAGGCCCTG 600  
 |||||  
 Db 587 GTGGTACCCTGTGGTGGAGGAGGATGCTTGTGGAATAGCAATTAACAATTAGGCCCTG 646  
 Qy 601 AAACCTAGTGTGAAGTGTATACGCTGTGAGCCCTCGAATGCGATGCTGCTACCTACCTGCT 660  
 |||||  
 Db 647 AAACCTAGTGTGAAGTGTATATGCTGNTGA-NCCTCAATGCGATGACTG-TACCAGTCC 704  
 Qy 661 AAACCTAGGAGGAGGAGTGAACCC 682  
 |||||  
 Db 705 AAGCTGAAGGGGACCTTATGCC 726

RESULT 11  
 AAA59295  
 ID AAA59295 standard; DNA; 608 BP.  
 XX  
 AC AAA59295;  
 XX  
 DT 07-NOV-2000 (first entry)  
 XX  
 DE N-terminal sequence of human serine racemase DNA.  
 XX  
 KW Serine racemase: N-methyl-D-aspartate receptor; neural death;  
 KW neural dysfunction; NMDA receptor; Parkinson's disease;  
 KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2000043526-A1.  
 XX  
 PD 27-JUL-2000.  
 XX  
 PF 18-JAN-2000; 2000WO-US00938.  
 XX  
 PR 19-JAN-1999; 99US-0116333.  
 PR 21-JUL-1999; 99US-0144839.  
 PR 28-JUL-1999; 99US-0145953.  
 XX





PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249220.  
PR 17-NOV-2000; 2000US-0249221.  
PR 17-NOV-2000; 2000US-0249222.  
PR 17-NOV-2000; 2000US-0249223.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249225.  
PR 17-NOV-2000; 2000US-0249226.  
PR 17-NOV-2000; 2000US-0249227.  
PR 17-NOV-2000; 2000US-0249228.  
PR 17-NOV-2000; 2000US-0249229.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX  
DR WPI; 2001-541565/60.  
XX  
XX  
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -  
XX  
XX  
PS Disclosure; SEQ ID NO 8040; 1701pp + Sequence Listing; English.  
XX  
XX  
CC The invention relates to novel genes (AB11004-ABA21534) and proteins  
CC (AB114678-AB118001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 20892 BP; 5768 A; 4125 C; 5164 G; 5835 T; 0 other;

Query Match 17.9%; Score 181.8; DB 22; Length 20892;  
Best Local Similarity 88.0%; Pred. No. 1.7e-46;  
Matches 198; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 593 AGGCCCTGAACCTAGTGTGAGGTATACGCTGCTGAGCCCTCGAATGCAGATGCT 652  
DB 13313 AGGCTCTGAACCTAGTGTGAGGTATATGCTGCTGAACTCAATGCAGATGCT 13254  
QY 653 ACCAGTCTAAACTGAAGAGGAAGTACCCCAATCTTCATCTCCAGAAACCATAGCAG 712  
DB 13253 ACCAGTCAAGCTGAAGGGAAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 13194  
QY 713 ATGGTGTCAAAATCCAGCATTCGCTTGAATACCTGGCCTATTATAGAGACCTTGTGATG 772

Db 13193 ATGGTGTCAAAATCCAGCATTCGCTTGAACACCTGCCTATTATCAGGACCTTGTGGATG 13134  
QY 773 ATGCTCTTCACTGTCTACCCGAAGATGAATCAAGTATGCAACCCAGC 817  
Db 13133 ATATCTTCACTGTCTACAGAGGATGAATTAAGGTGAGGCTCCAGC 13089  
RESULT 14  
AAC10869  
ID AAC10869 standard; cDNA: 861 BP.  
XX  
AC AAC10869;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein 5' EST, SEQ ID NO: 14944.  
XX  
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX  
OS Homo sapiens.  
XX  
PN EPI033401-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-0200610.  
XX  
PR 26-FEB-1999; 99US-0122487.  
XX  
PA (GEST ) GENSET.  
XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
XX WPI; 2000-500381/45.  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 1; SEQ ID 14944; 71pp + CD-ROM; English.  
XX  
CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.  
XX  
SQ Sequence 861 BP; 257 A; 167 C; 240 G; 191 T; 6 other;

Query Match 16.5%; Score 168.2; DB 21; Length 861;  
Best Local Similarity 88.8%; Pred. No. 6.6e-43;  
Matches 182; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 394 GACGAGTCCAGAGAAAGGTCACCTCAAGACATTATGCAAGAACAGAGGATCTTGTC 453  
Db 657 GAGATGTCCAGAGAAATGTTGCAAAAAGAGTTACAGAGAAACAGAGGATCATGTA 716  
QY 454 CATCCCAACCAAGGAGCCCTGAGTGTAGCTGGACAAAGGAACAATTCCTGGAAGTGTG 513  
Db 717 CATCCCAACCAAGGAGCCCTGAGTGTAGCTGGACAAAGGAACAATTCCTGGAAGTGTG 776  
QY 514 AACCAAGGTTCCCTTGGTGTAGTGCATGCTGTGTACCAGTAGGAGGAGGAATGTTGCT 573

```
Db 777 AACCAGGTTCTTGGTGGATGCACCTGGTGTACCTGTAGGTGGAGGAACTGCT 836
Oy 574 GGAATAGCATTACAAATTAAGGCC 598
Db 837 GGAATAGCATTACAGTTAAGGCTC 861

RESULT 15
AAV74370
ID AAV74370 standard; DNA; 31096 BP.
XX AC AAV74370;
XX 16-MAR-1999 (first entry)
XX Staphylococcus aureus contig SEQ ID #59.
XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
XX toxic shock syndrome; ds.
OS Staphylococcus aureus.
XX Key Location/Qualifiers
FH misc_feature 1201..1260
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 3001..3060
FT /*tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 4801..4860
FT /*tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 6601..6660
FT /*tag= d
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 8401..8460
FT /*tag= e
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 10201..10250
FT /*tag= f
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 12001..12060
FT /*tag= g
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 13801..13860
FT /*tag= h
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
```

```
FT misc_feature 15601..15660
FT /*tag= i
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 17401..17460
FT /*tag= j
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 19201..19260
FT /*tag= k
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 21001..21060
FT /*tag= l
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 22801..22860
FT /*tag= m
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 24601..24660
FT /*tag= n
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 26401..26460
FT /*tag= o
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 28201..28260
FT /*tag= p
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 30001..30060
FT /*tag= q
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
XX EP786519-A2.
XX 30-JUL-1997.
XX 07-JAN-1997; 97EP-0100117.
XX 05-JAN-1996; 96US-0009861.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
PI Rosen CA;
XX WPI; 1997-374922/35.
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus
FT stored on computer readable medium and used in the production of
```

anti-S.aureus vaccines

Claim 1; Page 452-469; 327lpp; English.

This sequence represents one of 5191 Staphylococcus aureus DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the S.aureus DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against S.aureus infection. The polypeptides can also be used in a kit for the immunodetection of S.aureus in a sample. S.aureus is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the S.aureus DNA sequences contained on the computer readable medium.

Sequence 31096 BP; 11857 A; 5243 C; 5477 G; 7488 T; 1031 other;

Query Match 13.7%; Score 139.8; DB 18; Length 31096;  
Best Local Similarity 48.5%; Pred. No. 5.4e-33;  
Matches 450; Conservative 0; Mismatches 468; Indels 10; Gaps 2;

Qy	35	TTGAAAAGCTCATATCAAGACTTCAAGACTTATCCACCTCACCCAGCTGTACAAAGCT 94
Db	1261	TCGAAGAAGCTTAAAGCAAGCAATTAACCATTTATTCGTGCAACACCTCTAATTAATCAAA 1320
Qy	95	CCATTTTGAATCAA---ATAGCAGCGGCGCAATCTTTTCTTCAAAATGTGAGCTCTCCAGA 151
Db	1321	TGTATTTAAGCCAAAGTATRACTAAAGGAATGTATTTCTAAATTTAGAAAATATGCAAT 1380
Qy	152	AAACTGGGTCTTTTAAAGATTCGAGGTGCCCTTAATGCCATCAGAGGCTTAATCCCTGACA 211
Db	1381	TCACAGGATCTTTTAAATTTAGAGCGCTAGCAAT-----NAAATTAATCACTTAAC 1433
Qy	212	CGCCAGAAGAGACCCCAAGCCGTAGTTACTACAGCAGCGGGAACCATGGCCAAAGCTC 271
Db	1434	AGATGAACAAAGAAAGAAAGGCATTTATCGCAGCATCTGCTGGGGAAGCCATGCACAAAGGTG 1493
Qy	272	TCACCTATGCTGTAAACTGGAAGGAATTCCTGCTTACATTTGTGGTCCCAACAGCTC 331
Db	1494	TTGCTTTAAGCAGCTAAATTTATAGGCATTTGATGCAACGATTTGTAATGCCCTGAACACGAC 1553
Qy	332	CCAACTGCAAGAACTGGCAATCCAGCCCTATGGAGCATCGATAGTATACTGTGACCCAA 391
Db	1554	CACAAGCGAAACAACAAGCAACAAAGGCTATGGGCAAGGTTATTTTAAAGGTAATAA 1613
Qy	392	GTGACGAGTCCAGAGAAAAGGTCACCTCAAGAAATTTATCGAAGAAACAAGAGGCATCTTGG 451
Db	1614	ACTTTAAGAACTAGACTTTTATATGGAAGAAATTTAGCGAAAGAAATGGCATGACAATCG 1673
Qy	452	TCCATCCCAACAGAGCGCTGCGATGATAGCTGGACAAGGAACAATTCGCTGGAAAGTGC 511
Db	1674	TTATCCATATGAGGATAAGTTTGTATGCGCAGCCCAAGGACAAATTTGGTTTAGAAATTT 1733
Qy	512	TGAACAGGTTCCCTTGGTATAGTCACTGGTGTGTGTTACAGTAGGAGGAGGAATGGTTG 571
Db	1734	TAGATGATATTTGGAATGTGAATACAGTCATCGTACCAGTTGGCGGTGGAGGATTAATTG 1793
Qy	572	CTGGAATAGCCATTAACAATTAAGCCCTGAAACCTAGTGTGAAGGTATACGCTCTGAGC 631
Db	1794	CAGTATTTGGCCGCGCATTAATAATCATTTAAACCTTCAATTCATATTTATCGGTGTTCAAT 1853
Qy	632	CCTGGAATGCGAGTACTGTCTACCAGTCTAAACTGAAAGGAGAACTGACCCCAATCTTC 691
Db	1854	CTGAAATGTTATGGTATGGCTGAGTCTTTCTATAGAGAGATTTAACTGAACATCGAG 1913

Qy	692	ATCCTCCAGAAACCATAGCAGATGGTGTCAAAATCCAGCATTTGGCTTGAATACCTGGGCTA 751
Db	1914	TGGATAGCACAAATAGCAGATGGTTGTGATGTAAAAAGTTCTCGTGGTGAACAAACATATGAAG 1973
Qy	752	TTATAAGAGACCTTTGTGATGATGCTCTTCACTGTCCACCGAAGATCAAAATCAAGTATGCAA 811
Db	1974	TAGTTAAACATTTTAGTAGATGAATTTATCTTGTACTGAAAGAGAAATTTGAACATGCTA 2033
Qy	812	CCCAGCTGGTGTGGGGGAGAATGAAACTGCCTATTGAGCCGACTGCTGGCGTGGCAGCTGG 871
Db	2034	TGAAGATTTAATCCAGCGTCCCAAAATTTATTACTGAAGGTGCAGGCGCATTTACCAACAG 2093
Qy	872	CTGCAGTCTGCTCAGCATTTTCCAAACAGTCTCTCCAGAAAGTAAAGAACGCTGTGCAATTG 931
Db	2094	CTGCAATTTTAAAGTGGAAAAATAAACAAATAAATGCTTGAAGATAAAAAATGTTGTTGCAT 2153
Qy	932	TACTCAGTGGGGGGAATCTAGACCTAAC 959
Db	2154	TAGTTTTCAGCGCGGAATGTTGACTTAAC 2181

Search completed: June 24, 2003, 04:26:07  
Job time : 291 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 03:22:18 ; Search time 186 Seconds  
(without alignments)  
8031.403 Million cell updates/sec

Title: us-09-889-609b-1

Perfect score: 1018

Sequence: 1 atgtgtgtcagctactgcatt.....ttaccagacggtctgttttaa 1018

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	143.4	14.1	31096	7	US-08-781-986A-59
2	87.4	8.6	1830121	9	US-10-329-960-1
3	77	7.6	954	10	US-09-974-300-4749
4	71	7.0	105184	9	US-09-847-513A-1
5	70.2	6.9	930	9	US-09-738-626-1091
6	70.2	6.9	3309400	9	US-09-738-626-1
7	60.6	6.0	1545	9	US-09-942-891-1
8	60.6	6.0	1545	9	US-09-942-891-5
9	60.6	6.0	1545	9	US-09-942-891-7
10	60.6	6.0	1545	9	US-09-942-891-8
11	55.2	5.4	1251	10	US-09-974-300-264
12	52.6	5.2	536165	9	US-09-939-964-1
13	46	4.5	926	10	US-09-974-300-4809
14	38.8	3.8	1379	9	US-10-098-841-197
15	38.4	3.8	15249	7	US-08-781-986A-102
16	37.2	3.7	1744	10	US-09-804-682-78
17	36.8	3.6	6320	9	US-10-037-270-135
18	35.2	3.5	1061	10	US-09-833-381-1809
19	34.6	3.4	31871	9	US-10-092-154-1403

20	34.6	3.4	31871	10	US-09-764-847-1403	Sequence 1403, Ap
21	34.6	3.4	44848	9	US-09-988-113-42	Sequence 42, Appl
22	34.6	3.4	44848	10	US-09-776-874A-42	Sequence 42, Appl
23	34.4	3.4	1049	9	US-10-123-155-358	Sequence 358, App
24	34.2	3.4	928	10	US-09-974-300-271	Sequence 271, App
25	34	3.3	486	9	US-09-918-995-32623	Sequence 32623, A
26	34	3.3	3655	10	US-09-903-187A-6	Sequence 6, Appli
27	34	3.3	3657	9	US-09-903-170C-6	Sequence 6, Appli
28	34	3.3	3657	10	US-09-903-180B-6	Sequence 6, Appli
29	34	3.3	3657	10	US-09-903-171A-6	Sequence 6, Appli
30	34	3.3	3657	10	US-09-903-188A-6	Sequence 6, Appli
31	34	3.3	3657	10	US-09-903-323A-6	Sequence 6, Appli
32	34	3.3	3657	10	US-09-903-325A-6	Sequence 6, Appli
33	33.8	3.3	498	10	US-09-864-761-32574	Sequence 32574, A
34	33.8	3.3	1274	9	US-09-983-802-75	Sequence 75, Appl
35	33.6	3.3	677	9	US-10-123-155-230	Sequence 230, App
36	33.4	3.3	1569	10	US-09-974-300-2651	Sequence 2651, Ap
37	33.4	3.3	2698	10	US-09-939-980-207	Sequence 207, App
38	33	3.2	1827	10	US-09-974-300-1928	Sequence 1928, Ap
39	33	3.2	6459	9	US-10-029-217A-6	Sequence 6, Appli
40	33	3.2	81940	9	US-09-759-508B-1	Sequence 1, Appli
41	32.8	3.2	552	9	US-10-123-155-196	Sequence 196, App
42	32.8	3.2	569	9	US-10-066-543-522	Sequence 522, App
43	32.8	3.2	1956	10	US-09-351-794A-1	Sequence 1, Appli
44	32.6	3.2	398	10	US-09-880-107-881	Sequence 881, App
45	32.6	3.2	763	9	US-10-153-668-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1  
US-08-781-986A-59  
Publication No. US20030054436A1  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,986A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31096 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

US-08-781-986A-59

Query Match		14.1%	Score 143.4;	DB 7;	Length 31096;
Best Local Similarity		48.6%	Pred. No. 1.1e-37;		
Matches 456;		Conservative 0;	Mismatches 472;	Indels 10;	Gaps 2;
QY	25	TTTGCTGATGTTCAAAAGCTCATATCAACATTCRAGACTCTATCCACCTCACCACGAGTG	84		
Db	1251	TTAGGAGATGATGAAGAGCTAAAGCAAGCATTAACACCATTTATTGTCGCAACACCTCTA	1310		
QY	85	CTAACAAAGCTCCATTTTGAATCAA---ATAGCAGGGCGCAATCTTTTCTTCAAAATGTGAG	141		
Db	1311	ATTAATCAATGATTTTAAAGCACAAGTATACTAAAGGAATGATTTCTAAATATAGAA	1370		
QY	142	CTCTCCAGAAAGCTGGGCTTTTAAAGATTCGAGTGGCCCTTAAATGCCATCAGAGCTTGA	201		
Db	1371	AATATGCAATTCACAGGATCTTTTAAATTTAGAGCGCTAGCAAT-----NAAATTA	1423		
QY	202	ATTCCTGACGCGCAGAGAGAGCCCAAGCGGTAGTTACTCAAGCAGCGGAACCAAT	261		
Db	1424	ATCACTTAACAGATGAACAAAAGAAAAGGCAATTATCGCAGCATCTGCTGGGGAACCAAT	1483		
QY	262	GGCCAAGCTCTCACCTATCTGCTAACTGGAAGGAATTCCTGCTTACATTTGTGGTTCC	321		
Db	1484	GCACAGGTGTGCTTTACACGCTAAATTTATAGGCATTTGATGCAACGATTTGATGCT	1543		
QY	322	CAAACAGCTCCCACTGCAAGAACTGGCAATCCAAGCCTATGGAGCATCGATAGTATAC	381		
Db	1544	GAACAGCACCACAGCGGAAACAACAAGCAACAAAAGGCTATGGGCAAGGTTATTTTA	1603		
QY	382	TGTGACCCCAATGACGAGTCCAGAGNAAGGTCACCAAGAAATATGCAAGAACAGAA	441		
Db	1604	AAAGSTAAAGCTTTTAAACGAACTAGACTTTATATGGAAGAATTAGCGAAAGAAATGCG	1663		
QY	442	GGCATCTTGGTCCATCCCAACAGAGCGCTGCAGTGATAGCTGGACAAGGAACATTCGC	501		
Db	1664	ATGACAACTGTCATCCATATGACGATAAGTTTGTAAATGCCAGGCCAAGAACATTTGT	1723		
QY	502	CTGGAAGTCTGAAACAGGTTCCCTTGGTAGATGCACCTGGTGTACCACTAGGAGGAGA	561		
Db	1724	TTAGAAATTTAGATGATATTTGGAATGTGAATACAGTCATCGTACCAGTTGGCGGTGA	1783		
QY	562	GGAATGGTTGCTGGATAGCCATTAACATTAAGGCCCTGAACCTAGTGTGAAGGTATAC	621		
Db	1784	GGATTAATTTGCAGGTATTTGCCACCGCATTTAAATCATTTTAAACCTTCAATTCATATTC	1843		
QY	622	GCTGCTGAGCCCTCGAATCGATGACTGCTACCAGTCTFAACTGAAAGGAGAAGTGAAC	681		
Db	1844	GGTGTCAATCTGAGAAATGTTTATGATGGCTGAGTCTTTCTATAAGAGAGATTTAACT	1903		
QY	682	CCCAATCTTCATCCTCCAGAAACCAATAGCAGATGGTGTCAAAATCCAGCAATTCGCTTGAAT	741		
Db	1904	GAACATCGAGTGGATAGCAACAATAGCAGATGGTGTGTGATGTAAAAGTTCCCTGGTGAACAA	1963		
QY	742	ACCTGGCTATTATAGAGACCTTTGGGATGATGCTTTCACCTGTCAACCAAGATGAATC	801		
Db	1964	ACATATGAAGTAGTTAAACATTTAGTAGATGAATTTATTTCTTTACTGAAGAAGAAAT	2023		
QY	802	AAATATGCAACCCAGCTGGTGGGGGAGAATGAAACTGCTCATTTGAGCCGACTGCTGCG	861		
Db	2024	GAACATGCTATGAAGATTTAATGCAGCGTGCANAAATTTACTCAAGGTCAGCGCA	2083		
QY	862	GTGGCACTGGCTGCAGTCTGCTCAGCAATTTCCAAACAGTCTCTCCAGAACTAAAGAAC	921		
Db	2084	TTACCAACAGCTGCATTTTAAAGTGGAAAATAAACAATAAATGGCTTGAAGATAAAAAT	2143		
QY	922	GTCTGCTATTGACTAGTGGGGGAATGTAGACCTAAC	959		
Db	2144	GTTGTTGCATTTAGTTTCAGCGGGGAATGTTGACTTAAC	2181		

RESULT 2  
US-10-329-960-1  
; Sequence 1, Application US/10329960  
; Publication No. US2003009277A1

GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, F  
; FILE OF INVENTION: Thereof, and Uses Thereof  
; FILE REFERENCE: PB186P1  
; CURRENT APPLICATION NUMBER: US/10/329,960  
; CURRENT FILING DATE: 2003-01-02  
; PRIOR APPLICATION NUMBER: US 09/643,990  
; PRIOR FILING DATE: 2000-08-23  
; PRIOR APPLICATION NUMBER: US 08/487,429  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/426,787  
; PRIOR FILING DATE: 1995-04-21  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1830121  
; TYPE: DNA  
; ORGANISM: Haemophilus influenzae  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (4747)..(4747)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (9921)..(9921)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (10150)..(10150)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (29298)..(29298)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36543)..(36543)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36551)..(36551)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36636)..(36636)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (40808)..(40810)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44416)..(44416)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44905)..(44905)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44975)..(44975)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (45593)..(45593)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (45732)..(45732)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature

```
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
```

```
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)
```

Query Match 8.6%; Score 87.4; DB 9; Length 1830121;

Best Local Similarity 45.7%; Pred. No. 3.9e-17;

Matches 304; Conservative 0; Mismatches 361; Indels 0; Gaps 0;

```
QY 215 CAGAGAGAGCCCAAGCGGTAGTTACTACAGACGCGGAACCACTGGCCCAAGCTCTCA 274
      ||||| ||| | ||||| | ||||| ||||| ||||| ||||| ||||| |||||
Db 794187 CAGACAAAAAGCAGCTGGCGTAATAGCAGCCCTCTCGCGGTAACTGCGCAAGCGGTGG 794246
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 275 CCTATGCTGCTAAACTGGAAGGAATTCCTGCTTACATTTGTTGCCCAACAGCTCCCA 334
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 794247 CATATCGCGGAACAAATTAGGCTTAAAGCATTAATTTATGCCACAAAACACCCCA 794306
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 335 ACTGCAAGAACTGGCAATCCCAAGCCTATGGAGCATCGATAGTATACTGTGACCCCAAGTG 394
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 794307 GCATTTAAAGTGGATGCAGTCGCGTGGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 794366
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 395 ACGAGTCCAGAGAAAGGTCACCTCAAGAATTTATGCAAGAAACAGAGGCAATCTTGGTCC 454
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 794367 TCGATGAAGCCAAAGCAAGGCAATCGAGCTTTCAAAGAAAAAAGACATGACATTTATTC 794426
```



[illegible]

## RESULTS

```

RES001.3
US-09-738-626-1091
; Sequence 1091, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIRO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1091
; LENGTH: 930
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1091

```

	Query Match	6.9%	Score 70.2;	DB 9;	Length 930;
	Best Local Similarity	47.2%;	Pred. No. 2.7e-13;		
	Matches 213;	Conservative 0;	Mismatches 238;	Indels 0;	Gaps 0;
Qy	131	TCAAAATGTGAGCTCTTCCAGAAAAC	TGGGCTCTTTTAAGATT	TGAGGTGCGCTTAATGCCA	190
Db	107	TCAAAAGCAGAGTTCTCTCCAAAG	TGCGCGTGTTCAAAACG	TGTGGAGCAATTCACCGCC	166
Qy	191	TCAGAGGCTTAATTCTGTACACG	CCAGGAAGCCCAAGCGGTAGTT	TACTCACA	250

Db	167	AGCTCGCAGCTTCCGAAAAACGGACTACTCGAACCCACAGGTTGGCATCGTCGCGCATCAG	226
Qy	251	GCGGAAACCATGGCCAAAGCTCTCACCTATGCTGCTAAACTGGAAAGGAATTCCTGCTTACA	310
Db	227	GCGGAAACGCAGGACTCGCAAATGCTTTGCCGCGAGCATCCTTAAAGCTTCCCGCCACGG	286
Qy	311	TTGTGGTTTCCCAAAACAGCTCCCAACTGCGAAGAAACTGGGCAATCCAAAGCCTATGAGGACAT	370
Db	287	TATTGGTGGCCGAAACTGCCCAAGTAAAAGTTGATCGCTCAAGCAATACGGTGCAA	346
Qy	371	CGATAGTATACTGTGACCCAGTGACAGTGCCAGAGAAAGGTGCATCAAGAATTTATGC	430
Db	347	CCGTGCAACAAATCCGATCTGAATATCGGAAGCATTTGAGCGAGCTCAAAACCTTTGAGT	406
Qy	431	AAGAAACAGAAAGGCATCTTGTTCCATCCCRAACAGAGCGCTGCAGTGATAGCTGGACAAG	490
Db	407	CGGAAATGGTGCTCTGTTTGGCCACCGCTACGACACCGCGACATCGCAGCTGGAGCAG	466
Qy	491	GAACAATTGCCCTGGGAAGTGTGAACAGGTTCCCTTTGTAGATGCACCTGGTGTTACCAAG	550
Db	467	GCGTCAITGGCGTAGAATATGTCGAAGATCTTCCCGACGTTGACACCATCGTGGTTGCTG	526
Qy	551	TAGAGGAGGAGGAATGGTTGCTGGAATAGC	581
Db	527	TCGGTGGCGGTGGACTCTATGACGAATCGC	557

## RESULT 6

```

US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738.626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

```

	Query Match	6.9%;	Score 70.2;	DB 9;	Length 3309400;
	Best Local Similarity	47.2%;	Pred. No. 5.2e-11;		
	Matches 213;	Conservative 0;	Mismatches 238;	Indels 0;	Gaps 0;
Qy	131	TCAATGTGAGCTCTTCCAGAAAACTGGGTCTTTTAAAGATTCGAGGTGCCCTTAAATGCCA	190		
Db	1039544	TCAAGCAGAGTTCCTCCAAAGTGC GGCGGTGTTCAAACGCGTGGAGCATTCACGCGC	1039485		
Qy	191	TCAGAGGCTTAATTCCTGCACAGCCAGAGAGAGCCCAAGCGCTAGTTACTCACAGCA	250		
Db	1039484	AGTTCGAGGTCGCGAAACAGGACTACTCCACCCAGCGTTGGCATCGTCGCGGCATCAG	1039425		
Qy	251	GCGGAAACCATGGCCAAAGCTCTACCTATGCTGCTAAACTGGGAAGGAATTCCTCGTTACA	310		



511	CTGAACCAAGTTCCCTTGGTAGATGCACCTGGTGTAACAGTACGAGAGAGAGAAATGGTT	570
Qy		
520	CTCCAGCAGGACGCCCATCTCGACGGCGATTTTGTGCCAGTCGCGCGCGCGCTCTGGCT	579
Db		
571	GCTGGAATAGCCATTACAATTAAAGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGA	629
Qy		
580	GCTTGGTGGCGGTGCTGATCAACAACAGTATGCGCAAAATCAAAGTGATGCGCGGTAGA	638
Db		

## RESULT 9

```

US-09-942-891-7
: Sequence 7, Application US/09942891
: Publication NO. US20030028917A1
: GENERAL INFORMATION:
: APPLICANT: Gruys, Kenneth James
: APPLICANT: Mitsky, Timothy Albert
: APPLICANT: Kishore, Ganesh Murthy
: APPLICANT: Slater, Steven Charles
: APPLICANT: Padgett, Stephen Rogers
: APPLICANT: Stark, David Martin
: TITLE OF INVENTION: Methods of Optimizing Substrate Pools and Biosynthesis of Poly-beta-hydroxybutyrate-co-poly-beta-hydroxyvalerate in Bacteria and Plants
: FILE REFERENCE: 11899.0155.DVUS02 (MOBT:155--3)
: CURRENT APPLICATION NUMBER: US/09/942,891
: CURRENT FILING DATE: 2002-02-01
: PRIOR APPLICATION NUMBER: US 09/313,123
: PRIOR FILING DATE: 1999-05-17
: PRIOR APPLICATION NUMBER: US 08/673,388
: PRIOR FILING DATE: 1996-06-28
: PRIOR APPLICATION NUMBER: US 08/628,039
: PRIOR FILING DATE: 1996-04-04
: PRIOR APPLICATION NUMBER: US 08/614,877
: PRIOR FILING DATE: 1996-03-13
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 7
: LENGTH: 1545
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic
US-09-942-891-7

```

```
Query Match      6.0%; Score 60.6; DB 9; Length 1545;
Best Local Similarity 46.5%; Pred. No. 7.9e-10;
Matches 195; Conservative 0; Mismatches 224; Indels 0; Gaps 0;
```

Qy		211	ACGCCAGAGAGAAGCCCAAGCGGTAGTTACTCACAGCAGCGAAACCATTGGCCAAGCT	270
Db		220	ACGGAAGAACAGAAAGCGCAGCGGCTCATCTCTTCTCGGGTAACACCGCAGGC	279
Qy		271	CTCACCTATCTGCTAAACTTGGRAGGAATTCCTGCTTACATTGTGGTTTCCCCAACAAGCT	330
Db		280	GTCGCGTTTCTTCTGCGCGGTTAGGCGTGAAAGCCCTGATCGTATTGCCAACCGCCACC	339
Qy		331	CCCAACTGCAAGAAACTTGGCAATCCAAAGCCTATGGAGCATCGATAGTATATACTGTGACCCA	390
Db		340	GCCGACATCAAGTCGACCGGCTCGCGCGGAGTGCTGCTCCAGCGCGG	399
Qy		391	AGTGACGAGTCCAGAGAAAAGGTACTCAAAGAAATTATGCAAGAAACAGAAAGGATCTTG	450
Db		400	AACCTTTGATGAAGCGAAACGCAAAAGCGATCGAACTGTACAGCAGCAGGGGTTCCACCTGG	459
Qy		451	GTCCATCCCAACGAGGACCTGCAGTATAGCTGGACAAGGAACAATTGCCCTGGAAAGTG	510
Db		460	GTCCGCGCGTGTGACCATCCGATGGTGATTGCCGGGCAAGGCACGCTGGCGCTGGAAGTG	519
Qy		511	CTGAACCAAGTGTCCCTTGGTAGATGCACCTGGTGTACCACTAGGAGGAGGAGGAATGTT	570
Db		520	CTCCAGCAGACGCCCATCTCGACCCGGTATTGTGCCAGTTCGCGCGCGCGGTCTGGCT	579
Qy		571	GCTGGATAGCCATTACAATTAAAGGCCCTGAAACCTAGTGTGAAGGTATACGCTGCTGA	629

## RESULT 11

US-09-974-300-264

[illegible]

## RESULT 10

```

US-09-942-891-8
; Sequence 8, Application US/09942891
; Publication NO. US20030028917A1
; GENERAL INFORMATION:
; APPLICANT: Gruys, Kenneth James
; APPLICANT: Mitsky, Timothy Albert
; APPLICANT: Kishore, Ganesh Murthy
; APPLICANT: Slater, Steven Charles
; APPLICANT: Padgett, Stephen Rogers
; APPLICANT: Stark, David Martin
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and Biosynthesis of Poly
; hydroxybutyrate-co-poly-beta-hydroxyvalerate in Bacteria and
; FILE REFERENCE: 11899.0155.DVUS02 (MOBT:155--3)
; CURRENT APPLICATION NUMBER: US/09/942,891
; CURRENT FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: US 09/313,123
; PRIOR FILING DATE: 1999-05-17
; PRIOR APPLICATION NUMBER: US 08/673,388
; PRIOR FILING DATE: 1996-06-28
; PRIOR APPLICATION NUMBER: US 08/628,039
; PRIOR FILING DATE: 1996-04-04
; PRIOR APPLICATION NUMBER: US 08/614,877
; PRIOR FILING DATE: 1996-03-13
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1545
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-942-891-8

```

Query Match	6.0%;	Score 60.6;	DB 9;	Length 1545;
Best Local Similarity	46.5%;	pred. NO. 7.9e-10;		
Matches 195; Conservative	0;	Mismatches 224;	Indels 0	

Qy	211	ACGCCAGAAGAGAACCCAAAGCCGTAGTTACTCACAGCAGCGGAACCATTGCCCAGCT	270
Db	220	ACGGAAGAACAAAGAGCAGCGGTGATCACTGCTTCTCGGGTAACCAACCGCAGGGC	279
Qy	271	CTCACCTATGCTGCTAACTGGAAGGAATTCCTGCTTACATGTGTGGTTCCTCCCAACAGCT	330
Db	280	GTCGCGTTTCTCTCGCGGTTTAGGCCTGAAGGCCCTGATCGTTATGCCAACC GCCCAC	339
Qy	331	CCCAACTGCAGAAACTGGCAATCCAAGCCTATGGAGCATCGATAGTATAC'TGTACCCA	390
Db	340	GCCGACATCAAAGTCGACCGGCTGCGCGCTTCGGCGGAGTGCTGCTCACGGCGG	399
Qy	391	AGTGACGAGTCCAGAGAAAAGTCACTCAAAGAA'TTATGCAAGAAACAAGAGGCATCTTG	450
Db	400	AAC'TTTGTGAAGCGAAACGCAAGCGATCGAACTGTCAACAGCAGGGGTTCACTCGG	459
Qy	451	GTCCATCCCAACACGAGGACCTGCAGTGATAGCTGCACAAGGAACAATTCGCCCTGGAAGTG	510
Db	460	GTGCGCGGTTGACCATCCGATGTGTGATTCGGGGCAAGGACCCTGCGCTGTGAAC'GTG	519
Qy	511	CTGAACACAGGTTCCTCTTGGT'AGATGCAC'TGGTGGTACCAGTAGGAGGAGGAATGGTT	570
Db	520	CTCCAGCAGGACGCCATCTCGACCGCGTATTTGTGCCAGTCGGCGCGCGCTGGCT	579
Qy	571	GCTGGATATAGCCATTACAATTAAGGCCCTCGAACTAGTGTGAAGGATATACGCTGCTGA	629
Db	580	GTTTCGTCGGCGGTGCTGATCAAACTCACTGATGCGCGCAATCAAAAGTATGATCCCGT	638

## RESULT 11

US-09-974-300-264

```
; Sequence 264, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; US-09-974-300-264

Query Match          5.4%; Score 55.2; DB 10; Length 1251;
Best Local Similarity 44.9%; Pred. No. 5.2e-08;
Matches 354; Conservative 0; Mismatches 413; Indels 21; Gaps 3;

QY 18 CATCTCTTGTGCTGATGTTGAAGAGCTCATATCAACATTCAGAGCTCTATCCACCTCAC 77
Db 14 CATCCAAGTCAAGACATTTTAAAGGCCACCAAAATGTAAGAGATGTTGTATTTCATAC 73
QY 78 CCCAGTGCCTAACAGCTCCATTTTGAATCAAAATAGCAGGCGCAATCTTTTCTTCAAAATG 137
Db 74 CCCTCTCCAAAATAATGAGAGCTGTCGAGAGGTACGATTGCAAGCTATATTAAAG 133
QY 138 TGAGTCTTCCAGAAACTGGTCTTTTAAAGATTCGAGGTGCCCTTAATGCCATCAGAGG 197
Db 134 AAGAGACCTCGCAGTGTGTCGGCTCTTTTAACTGAGAGGCGCTATTAT-----AA 184
QY 198 CTTAATTTCTGCACGCCAGAGAGAGAGCCCAAGCCGTAGTTACTCAGAGCAGCGGAAA 257
Db 185 ATTGAAGCAGCTTTCAAAAGAAACGACGAAACGGGGTGTGCGCCAGTGCAGGAAA 244
QY 258 CCATGCCCAAGCTCTCACCTATGCTGCTAACTTGAAGAAATTCCTGCTTACATTTGGT 317
Db 245 CCATGCCGAGGAGTGCCTTTTCCCTGCAACATCTCGGTATTACCGGAAAGATTTTAT 304
QY 318 TCCCCAAACAGCTCCCAACTGCAGAAACTGCGAATCCAGCCTATGGAGCATCGATAGT 377
Db 305 GCGGTGCGAGCAGCCCGAGGCAAAAATTTCCCAAGTCSAATTTGTCGCAAGGAATACGT 364
QY 378 ATACTGTGACCCCAAGTGCAGAGTCCAGAGAAAGGTCACTCAAGAAATATG----- 429
Db 365 CGAGATCATCTGCACAGGCGACACGTTTGATGTTTATCAAGGCGAGTGCCTGCTG 424
QY 430 -CAAGAAACAGAGGAGCATCTTGCTCCATCCCAACAGGAGCTGCAGTATAGCTGGACA 488
Db 425 CGAGGAAGAAAAGCGCGCATTTATCCACCGCTTGACGATCCTCGCGTATGCTGCGCCA 484
QY 489 AGGAACAATTTGCCCTGGAAGTGTGAACCAAGTTCCTTGGTA---GATGCACTGGTGT 545
Db 485 GGGTACGGTGGCGGTGCAATCTTAAACGATATCGAAACGGAACCCCATTTATTATCGC 544
QY 546 ACCAGTAGGAGGAGGAGGAATGTTGCTGGAATAGGCATTTACAAATTAAGCCCTGAACC 605
Db 545 AAGCGTAGGAGGAGGGGTCTGCTTTCAGAGTGGGAACCTATATGAAAACGCTGTCCCC 604
QY 606 TAGTGTGAAGGTATAGCTGCTGAGCCCTCGAATGCAGATGACTGCTACCACTTAACCT 665
Db 605 GGATACAAAGCTGATCGCGGTGAGCCTAAAGCGCGCGCTCTTTTGTGAATCAACAA 664
QY 666 GAAGAGAGAACTGACCCCAATCTTCATCCTCCAGAAACCATAGCAGATGGTGTCAATC 725
Db 665 AAAAGCGAAGTGCATCTCTTGAGAAAATCGATAAAATTTGTCGATGGAGCGCTGTGCA 724
```

```
QY 726 CAGCATTTGCTTGAATACCTGCGCTATTATTAAGAGACCTTGTGGATGATGCTTCACTGT 785
Db 725 GAAATCGCGGAGAGACATTTAAACGCTTGAACCGTTTGCATGATATTTTGCTCGT 784
QY 786 CACCGAAG 793
Db 785 TCCTGAAG 792

RESULT 12
US-09-939-964-1
; Sequence 1, Application US/09939964
; Publication No. US20030054522A1
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
; TITLE OF INVENTION: Plasmid
; FILE REFERENCE: CARP0088
; CURRENT APPLICATION NUMBER: US/09/939,964
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/214,808
; PRIOR FILING DATE: 1999-06-22
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium
; US-09-939-964-1

Query Match          5.2%; Score 52.6; DB 9; Length 536165;
Best Local Similarity 44.5%; Pred. No. 2e-05;
Matches 263; Conservative 0; Mismatches 319; Indels 9; Gaps 1;

QY 35 TTGAAAAGCTCATATCAACATTCAGACTCTATCCACCTCACCCAGTGCCTAACAGCT 94
Db 410024 TCAGGGGCGCGTGAGCGGATCGAAGAGCAGCTCTTCGCTACCCCTCTGAGACATCAA 410083
QY 95 CCATTTTGAATCAATAGCAGGCGCAATCTTTTCTTCAAAATGTGAGCTCTTCCAGAAA 154
Db 410084 GGTGCGTAACCGAACTTACCGGAACCTCAGCTCAGCTCAGACACTATCAGCGCA 410143
QY 155 CTGGGTCTTTTAAGATTCAGGTGCCCTTAATGCCATCAGAGGCTTAATCTCGACAGC 214
Db 410144 CTGGTAGCTTTAAGCTGCGGGTGCACAAACGCAA-----TTCTTTCAACTCAGCC 410194
QY 215 CAGAAGAGAACGCCAAAGCCGTAGTTACTCAGAGCAGCGGAACCATGGCCAGCTCTCA 274
Db 410195 CGTCGATCGGCACGTGGGGTTATTGCGGATCTACGGGCAATCAGGACGGCTCTTT 410254
QY 275 CCTATGCTCTAACTGGAAGGAATTCCTGCTTACATTTGTTTCCCAAAACAGCTCCCA 334
Db 410255 CCTACGCCCAAAAGCGTGGCTCTCGCGCCACCATCTGCGATCTGCGATCTTGTTCAG 410314
QY 335 ACTGCAAGAACTGGCAATCCAAAGCCTATGGAGATCGATAGTACTGTGACCCCAAGTG 394
Db 410315 AAAACAAGGTTTTCCGAGATCCGGAAGCTTTGGCGCACAGTTCGGATAGTGGATCGTCAC 410374
QY 395 ACGAGTCCAGAGAAAAGGTCACTCAAGAAATTTATCAAGAAACAGAACGCTCTGGTCC 454
Db 410375 AAGCATGCGCAAGTCGGAAGTCGAGCGCTCGTCCGGAGGAAGGCTCAGCATGATCC 410434
QY 455 ATCCCAACAGGAGGCTGCAGTGTAGCTGGCAAGAAACAATTCCTCGAAGTGCCTGA 514
Db 410435 CGCCTTTTCATCACCAGCATATCATCGCGCGCAGCAACCGCTCGCTCTTTGAGATCGTTG 410494
QY 515 ACCAGTTCCTTGGTAGATGCATGCTGGTGTACCAAGTAGGAGGAGGAGGAATGGTGTG 574
Db 410495 AGCGATGCGGACGTCGCGATGTTCTTCCACTGTCGGTGGCGCTTTGGTGCAG 410554
```

```
QY 575 GAATAGCCATTACAATTAAGCCCTGAACCTAGTGTGAAGGTATACGCTG 625
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410555 GCCTGCAGCAGCGGTGAAGCCACTGGCGCTCATCGCAGGATCATCGGTG 410605

RESULT 13
US-09-974-300-4809
; Sequence 4809, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: Expression
; CURRENT APPLICATION NUMBER: 10085.500-US
; CURRENT FILING DATE: 2001-10-05
; PRIOR FILING DATE: 2000-10-06
; PRIOR FILING DATE: 2000-10-06
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4809
; LENGTH: 926
; TYPE: DNA
; ORGANISM: Bacillus clausii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(926)
; OTHER INFORMATION: n = A,T,C or G
US-09-974-300-4809

Query Match 4.5% Score 46; DB 10; Length 926;
Best Local Similarity 44.2%; Pred. No. 6.6e-05;
Matches 346; Conservative 0; Mismatches 415; Indels 21; Gaps 3;

QY 31 GATGTTGAAAAGCTCATATCAACATTCAGACTCTATCCACCTCAGCCAGTGCTAACA 90
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 GATATTATTCGCAAAATCAACAAATTAAGATGTGTCGACGATACACCTTTGCAAAA 65

QY 91 AGCTCATTTTGAATCAATAGCAGCGGCAATCTTTTCTTCAATGTAGCTCTCCAG 150
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 GATCAGGTTTGTCTGAACGATATGATGCACAGTTTACTTAAACGGGAAGACTTGCAA 125

QY 151 AAAACTGGGCTTTTAAAGATTTCGAGTGGCCCTTAATGCCATCAGAGCTTAATTCCTGAC 210
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 GTTGTGAGATCTTTTAAATTCGGCGCTTATACCAATTTCTCGCTAT----- 177

QY 211 ACGCCAGAGAGAGAGCCCAAGCCGTAGTTACTCAGCAGCGGAAACCATGGCCAAAGCT 270
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 -CAAGAAGAACTTGCCTGCGTGTATGCGCAAGCGCTGGAACCATGCACAAAGG 236

QY 271 CTCACCTATGCTCTAACTGGAAGNAATCTCTTACATTTGTTGGTTCCTCCCAAGCT 330
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 GTTGCCTATTCATGCGCGCGCTTTAAAGTCAAAAGGTGTTATTTTCATGCGGACTACACG 296

QY 331 CCCAACTGCAAGAACTGGCAATCCAGCTATGAGCATCGATAGTATATCTGACCCA 390
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 CCAAGCAAAAGTCGCACAAAGTGAATTTTGGCAGGGACTATGTAGACGTACGGTTA 356

QY 391 AGTGACGAGTCCAGAGAAA-----GGTCACATCAAGAAATTTGCAAGAAACAGAA 441
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 ATTGGCGATACGTTTGATGATTCTTATGCGGCGCCATTCATATTGCGAAGAACAGGAA 416

QY 442 GGCATCTTGTCATCCCAACAGGAGCCCTGAGTATAGTGCAGCAAGAACAATGCC 501
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 ATGACGTTTCATCCATCCATTTTAAACAGACAAAGTATTTGCCGGGGAAGCAAGTCGGA 476

QY 502 CTGAAGAGTGTGA---ACCAGGTTCCCTTGGTAGATGCACCTGGTGGTACAGTAGGAGGA 558
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 CTTGAATTAATGACATCGAGGAACCCAGATTTTGTGTTTTCATCAATCGCGGC 536
```

```
QY 559 GGAGGAATGTTGCTGGAATAGCCATTACAATTAAGCCCTGAACCTAGTGTGAAGGTA 618
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 GGTGGGCTTATTAGTGGGATGGCCACCTATATAAAAAGCGTTACGCCCAACTACGAAATG 596

QY 619 TACGCTCTCAGCCCTCGAATGAGATGACTGTCTACCAGTCTAAACTGAAGAGAACTG 678
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 597 ATTGGTTGTAACCTGCTGTGCTGCATCATGACGGAATCAATTAAGCAAGGCAACGTT 656

QY 679 ACCCCCAATCTTCATCTCCAGAAACCATAGCAGATGGTGCAAAATCCAGCATTTGGCTTG 738
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 657 GTTGAGTTAGACGAATTCGAAATTTGTGATGGAGCGGCTGTGAAAAAAGTCGGCGAC 716

QY 739 AATACCTGGCTATTATAAGAGACCTTTGTGGATGATGCTTCTCACTGCACCGAAGATGAA 798
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 717 AAAACACTGCAAAATTTGCAAAAGAACTGCTAGATGACATTTATTGTTGTGCCCAAGAGGCAAA 776

QY 799 AT 800
  ||
Db 777 AT 778
  ||

RESULT 14
US-10-098-841-197
; Sequence 197, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Felyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 784CIP2
; CURRENT APPLICATION NUMBER: US/10/098,841
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 197
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (182)..(1171)
US-10-098-841-197

Query Match 3.88; Score 38.8; DB 9; Length 1379;
Best Local Similarity 49.0%; Pred. No. 0.027;
Matches 143; Conservative 0; Mismatches 137; Indels 12; Gaps 1;

QY 42 AGCTCATATCAACATTCAGACTCTATCCACCTCAGCCCAAGTGTAAACAGCTCAATTTT 101
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 AGAGCATGCCAAGCAGGAGGCCCTTTCACGTGGTCACACCTCTGTGTGGAGAGCTGGCGCT 258
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:11:01 ; Search time 276.093 Seconds  
(without alignments)  
132.862 Million cell updates/sec

Title: US-09-889-609B-8

Perfect score: 1740

Sequence: 1 MCAQYCISFADVEKAHINQ.....TSLNVGQARPAPIQTVSV 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 41779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 41779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB\_PEP.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB\_PEP.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB\_PEP.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB\_PEP.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB\_PEP.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB\_PEP.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB\_PEP.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB\_PEP.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB\_PEP.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB\_PEP.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB\_PEP.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB\_PEP.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB\_PEP.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB\_PEP.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	357.5	20.5	310	9	US-09-738-626-4591
2	307	17.6	436	9	US-09-738-626-5828
3	201.5	11.6	306	10	US-09-815-242-13657
4	196	11.3	308	9	US-09-738-626-6317
5	177	10.2	325	9	US-09-931-457A-32
6	175.5	10.1	310	10	US-09-815-242-10707
7	168	9.7	323	10	US-09-815-242-10239
8	167	9.6	383	9	US-09-931-457A-65
9	161	9.3	323	10	US-09-815-242-13862
10	160.5	9.2	325	9	US-09-931-457A-31
11	160.5	9.2	386	9	US-09-931-457A-66
12	160	9.2	398	9	US-09-931-457A-62
13	159.5	9.2	313	10	US-09-815-242-12146
14	159.5	9.2	313	10	US-09-815-242-12800
15	157.5	9.1	303	10	US-09-815-242-5437
16	157	9.0	415	9	US-09-931-457A-64
17	156.5	9.0	324	9	US-09-905-290A-4
18	154.5	8.9	324	10	US-09-815-242-11879
19	143.5	8.2	316	10	US-09-815-242-11179

20	141.5	8.1	314	10	US-09-815-242-5737	Sequence 5737, Ap
21	129	7.4	306	10	US-09-815-242-11292	Sequence 11292, A
22	115.5	6.6	317	9	US-09-738-626-5838	Sequence 5838, Ap
23	111.5	6.4	250	10	US-09-815-242-11635	Sequence 11635, A
24	110.5	6.4	853	10	US-09-815-242-11363	Sequence 11363, A
25	110.5	6.4	26926	9	US-09-739-508B-2	Sequence 2, Appli
26	109	6.3	1187	10	US-09-935-291A-4	Sequence 4, Appli
27	105.5	6.1	852	10	US-09-815-242-11535	Sequence 11535, A
28	102	5.9	164	10	US-09-734-017A-28	Sequence 28, Appl
29	94.5	5.4	2382	9	US-10-196-935A-2	Sequence 2, Appli
30	93.5	5.4	540	9	US-09-742-096-22	Sequence 22, Appl
31	93.5	5.4	1786	9	US-09-742-096-3	Sequence 3, Appli
32	92.5	5.3	202	10	US-09-734-017A-24	Sequence 24, Appl
33	90.5	5.2	212	9	US-09-742-096-25	Sequence 25, Appl
34	90.5	5.2	442	9	US-10-117-417-6	Sequence 6, Appli
35	90.5	5.2	442	9	US-10-117-417-16	Sequence 16, Appl
36	90.5	5.2	630	9	US-09-742-096-5	Sequence 5, Appli
37	89.5	5.1	297	9	US-09-941-947A-20	Sequence 20, Appl
38	89.5	5.1	297	10	US-09-934-903-14	Sequence 14, Appl
39	89.5	5.1	297	10	US-09-934-868-72	Sequence 72, Appl
40	88.5	5.1	259	9	US-09-996-634-133	Sequence 133, App
41	88.5	5.1	259	9	US-09-997-181-133	Sequence 133, App
42	88.5	5.1	259	9	US-09-997-182-133	Sequence 133, App
43	88	5.1	947	9	US-10-101-464A-73	Sequence 73, Appl
44	88	5.1	1493	10	US-09-858-754-4	Sequence 4, Appli
45	88	5.1	1642	10	US-09-925-442-2	Sequence 2, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-738-626-4591  
; Sequence 4591, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OKAZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4591  
; LENGTH: 310  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4591

Query Match 20.5%; Score 357.5; DB 9; Length 310;

Best Local Similarity 30.3%; Pred. No. 2.5e-25;

Matches 101; Conservative 52; Mismatches 139; Indels 41; Gaps 9;

Oy 7 ISFADVEKAHINQDSIHLTVLTSLNLTAGRNLFKCELFKTSFKIRGALN--- 62

Db 2 LTLDNDVITAQORTAPHVRRTPLEFA---DPIDGTQIWIKAEFLOKCGVFKTRGFNROLA 58

Oy 63 -AIRGLPDTPEPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPOTAPNCKKIAIOAY 121

-Db 59 ASENGLLDPT-----VGIVAASGAGNAGLAFASLSVPATVLVPEAPQVVKVDRLKQY 113  
QY 122 GASIVYCDPSDESREKVTQIMQETEGILVHPNQEPFAVITAGQGTIALEVLNQVPLVDALV 181  
Db 114 GATVQOIGSYBAFAFAAQTFFSETGALFCHAYDQPDIAAGAGVIGLEIVEDLPVDTIV 173  
QY 182 VPVGGGOWAGIATITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVK 241  
Db 174 VAVGGGLYAGIAAVAA--HDIKVAVPEPSKPTTLHNSLIAGQ-----PVDVNVSGIA 225  
QY 242 S-SIG-----LNTWPIIRLDVDDVFTVTEDEIKYATOLVWGRMKLLIEPTAGVA 289  
Db 226 ADSIGARQIGREAFDIATAHPPIGVLVD-----EALIAARRHLWDNVRIPAEGHCAAAA 279  
QY 290 LAAVLSQHQTQVSPVKNVCIVLSGGNVDLTSL 322  
Db 280 LASLTGAYKPAADE--KVAVIVCGANTDLTTL 310

## RESULT 2

US-09-738-626-5828  
; Sequence 5828, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 5828  
; LENGTH: 436  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5828

Query Match 17.6%; Score 307; DB 9; Length 436;  
Best Local Similarity 28.1%; Pred. No. 2,2e-20;  
Matches 94; Conservative 58; Mismatches 162; Indels 20; Gaps 9;  
QY 7 ISFADVEKAHINQDSIHILTPVLTSILNQIAGRNLFKCELFQKTSFKIRGALNAIRG 66  
Db 21 IRAADLTQARISSVIATPQYCPRLSEETGAELIYLRKEDLQDVRYSKIRGALNSGAQ 80  
QY 67 LIPDTPPEPKAVKAVYHSSNGHGAALTYAAKLEGIPAYIVVPTAPNCKKLAIOAYG---A 123  
Db 81 L---TQQRDAGIVAASAGNHAGVAYVCKSLGVQGRIVVPVTPQKQRDRIMVHGFEV 137  
QY 124 SIYVCDPSDESREKVTQIMQETEGILVHPNQEPFAVITAGQGTIALEVLNQVPLV----DA 179  
Db 138 LSVVTTGNFDEASAAHAEDEARTGATLIEFPDARNTVIGQTVAAAILSQLTSMGRKSADH 197  
QY 180 LVVPVGGGOWAGIATITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADG 239  
Db 198 VWVPVGGGGLLAGVSYMADMAPRTAIVGIEPAGAAS--MQAALHNGGPITLETVDVFDVG 256

QY 240 VK--SSIGLNTWPIIRLDVDDVFTVTEDEIKYATOL--VWGRMKLLIEPTAGVALAAVLSQ 296  
Db 257 AAVKRVGDLNLYTIVENKQGRVHMSATEGAVCTEMLDLYQNEGIIIEAPAGALSIALKEM 316  
QY 297 HFQTVSPVKNVCIVLSGGNVDLTSLNWMYQGAER 330  
Db 317 SF---APGSVVVVCII--SGGNNDV--LRYAEIAER 344

## RESULT 3

US-09-815-242-13657  
; Sequence 13657, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011a  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13657  
; LENGTH: 306  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-815-242-13657

Query Match 11.6%; Score 201.5; DB 10; Length 306;  
Best Local Similarity 25.3%; Pred. No. 1e-10;  
Matches 74; Conservative 53; Mismatches 124; Indels 41; Gaps 9;  
QY 18 NTODSIHLTPVLTSILNQIAGRNLFKCELFQKTSFKIRGALNAIRGLIPDTPPEPK 77  
Db 6 NITELIGQTPVKNLVNIVPEGAADVYKLEAFNPSSVKDRIALSMIEKAEQDILKPGS 65  
QY 78 AVYTHSSGNHGQALTYAAKLEGIPAYIVVPTAPNCKKLAIOAYGASIVYCDPSDESREK 137  
Db 66 TIVEATSGNTGTGLSWVGAAGKYKVIVMPETMSVERRKIIQAYGAELVLTGSEGMKA 125  
QY 138 V--TQRMQETEGILVHPNQEP-----VIAQGQTIALEVNLQVPLVDALVVP 183  
Db 126 IAKAQEIAAERDGLPLQFDNPNPEVHERTTGAELIAAFKDGDL-----DAFVAG 176  
QY 184 VGGGOWAGIATITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADG--VKS 242  
Db 177 VTGTTGTSVSHALKSENSNIQFAVE---ADE--SAILSGE--KPGPHKIQGISAGFIPD 230  
QY 243 SIGLNTWPIIRLDVDDVFTVTEDEIKYATOLVWGRMKLLIEPTAGVALAAVL 294  
Db 231 TLDTRAY-----DGIVRVTSDDALALREIGGEGFLV-----GISSAAAI 271

## RESULT 4

US-09-738-626-6317  
Sequence 6317, Application US/09738626  
Publication No. US20020197605A1

## GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO

APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIALI, KEIKO

APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAKO

APPLICANT: SENOH, AKIHIRO

APPLICANT: IKEDA, MASATO

APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-125

CURRENT APPLICATION NUMBER: US/09/738,626

PRIOR FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: JP 99/377484

PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988

PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059

SOFTWARE: PatentIn ver. 3.0

SEQ ID NO 6317

LENGTH: 308

TYPE: PRT

ORGANISM: Corynebacterium glutamicum

US-09-738-626-6317

Query Match 11.3%; Score 196; DB 9; Length 308;

Best Local Similarity 28.4%; Pred. No. 3.4e-10; Mismatches 109; Indels 58; Gaps 12;

Matches 80; Conservative 35; Mismatches 109; Indels 58; Gaps 12;

QY 18 NIQDSIHLTPVLTSSILNOIAG---RNLFKCELFQKTSFKIRGALNAINRGLDPTPEE---74

DB 4 NITETIGHTPLVNLKLTGLDNLVLFKESFNANPVKDRIGL---AIVDAEAKSGAL 59

QY 75 KP-KAVVTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIOAGSIVYC-----128

DB 60 KPGGTIVATSGNTGIGLAFIAAAGYRLIICMPASMSLERRTILRAFGAELVLT 119

QY 129 --DPDESREKVTQIMQETEGTLVHPNOEPVIAQ-----GTIALEVLNQVP 175

DB 120 MQAKDKADEIVAER-----ENAVLARQFENEANPRVHRDITAKEILEDYD 165

QY 176 -LVDAIVPVVGGGVMVAGIATIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLRHPE 234

DB 166 GNVDIFVAFSGTGTGTGVQVLKKNADQVTVTPE-----ASPL---LTAGKAGPH 216

QY 235 TIADGVKSSIGLNTWPII--RDVDDVFTVT--EDEIKYATQL 273

DB 217 KI-----QGIGANFIPEVLDRKVLDDVLTVSNEDALAFSRKL 253

RESULT 5

US-09-931-457A-32

Sequence 32, Application US/09931457A

Patent No. US20020157132A1

GENERAL INFORMATION:

APPLICANT: Falco, S. Carl

APPLICANT: Allen, Stephen M.

TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes

FILE REFERENCE: BB1116 US CIP

CURRENT APPLICATION NUMBER: US/09/931,457A

CURRENT FILING DATE: 2002-02-22

PRIOR APPLICATION NUMBER: 09/424,976

PRIOR FILING DATE: 1999-12-02

PRIOR APPLICATION NUMBER: 60/065,385

PRIOR FILING DATE: 1997-11-12

PRIOR APPLICATION NUMBER: 60/049,406

PRIOR FILING DATE: 1997-06-12

NUMBER OF SEQ ID NOS: 72

SOFTWARE: Microsoft Office 97

SEQ ID NO 32

LENGTH: 325

TYPE: PRT

ORGANISM: Citrullus lanatus

US-09-931-457A-32

Query Match 10.2%; Score 177; DB 9; Length 325;

Best Local Similarity 23.9%; Pred. No. 2.2e-08;

Matches 74; Conservative 53; Mismatches 112; Indels 70; Gaps 12;

QY 18 NIQDSIHLTPVLTSSILNOIAG---RNLFKCELFQKTSFKIRGALNAINRGLIP 69

DB 11 DVTELTGNTPLV---YLNRVDDCVARVAAKLEMPEPCSSVKDRIGYSMLSDAENKGLI- 66

QY 70 DTPEEKPAVVTTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIOAGSIVYCD 129

DB 67 -TPGE--SVLIEPTSGNTGIGLAFIAAAGYRLIICMPASMSLERRTILRAFGAELVLT 123

QY 130 P-----SDSREK-----VTQIMQETEGTLVHPNOEPVIAQGTI 166

DB 124 PARGMKGAVQKAEIEKAKTPNSYILOQFENPANPKIHYETTG-----PEIWRGSG-- 173

QY 167 ALEVLNQVPLVDALVVPVGGGVMVAGIATIKALKPSVKVYAAEPSNADDCYQSKLKGEL 226

DB 174 -----GKIDALVSGTGTGTGVQVLKKNADQVTVTPE-----VSAIILSGG 219

QY 227 TPNLHPPETIADG-VKSSIGLNTWPII--RDVDDVFTVT--EDEIKYATQLVWGRMKLLIETP 285

DB 220 KPGPHKIQIGTAGFIPGVLDVN-----LLDEVIQVSSSESIETAKLLAKKGLLVGIS 272

QY 286 AGVALAAVL 294

DB 273 SGAAAAAAAI 281

RESULT 6

US-09-815-242-10707

Sequence 10707, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

Query Match 9.7% Score 168 DB 10 Length 323

Query Match	9.6%	Score 167;	DB 9;	Length 383;
Best local Similarity	23.7%;	Pred. No. 2.4e-07;		
Matches	77;	Conservative 58;	Mismatches 132;	Indels 58;
Gaps	14			

  

QY	6	CISFADYEKAHIN	QDSIHLPVLTSS	TLN-----QIAGR-----	NLFF	44
Db	38	CNPFKKVSSPITCK	AVLSPPSTIEGLN	IAEDVSQIGIKTPM	VLNNVSKGS	VANIAA 96
QY	45	KCELFORTGSPKIR	GAUNAINGLIPD	TPEEK-----PKAVVTHSS	NGHQALTYAA	KL 98
Db	97	KLESMPECCSVK	DRIGYSMI-----	DDAEQKGVITPGK	TTLVEPTSGT	GLCLAFIAAAR 151
QY	99	GIPAYIVVPOTAP	NCKLIAQVAGSIV	CDPDSER--	EKVQRIQOE	TGILLVHPNQ 155
Db	152	GKYITLTPASMS	ERRVILKAFCAEL	VLTPDPAKGMKGA	VEKAEBELK	KTPDPSYMLQOFD 211
QY	156	EPA-VIAGQGTIA	LEVLNQVP-LYD	ALVVPVGGGGMV	AGIAITIKAL	PKSVKVVAAEPSN 213
Db	212	NPANPKTHYETGP	IEWDTGKQVDIF	VAGIGTGTGCT	ISCGVRYL	KERNPGVQVIGIEPT- 270

QY 214 ADCVQSKLKGELTPNHLHPETIADG-VKSSIGLNTWPIIRDLVDDVFTVTDE-IKYAT 271  
Db 271 -----ESNLSGCKGPHIKIQIGLGFVPSNLDLG-----VMDEVIESSEAEVEMAK 318  
QY 272 QLVWGRMK--LLIEPTAGVALAVAL 294  
Db 319 QLA---MKEGLLVGISSGAAAAAVAL 340

## RESULT 9

US-09-815-242-13862  
; Sequence 13862, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: fastseq for windows Version 4.0  
; SEQ ID NO 13862  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Salmonella typhi  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(323)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-815-242-13862

Query Match 9.3%; Score 161; DB 10; Length 323;

Best Local Similarity 24.1%; Pred. No. 6.9e-07;  
Matches 74; Conservative 55; Mismatches 136; Indels 42; Gaps 11;

QY 26 TPVLTSSILNOIAGRNLFKCELFQKTSFKIRGALNAI-----RGLIPDTPEEKPKA-V 79  
Db 15 TPLVR---LNRINGRILAKVESRNFVSFKRIGANMTWDAEKRGL-----KPGVEL 65  
QY 80 VTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIOAYGASTIVCDPSDESREKVT 139  
Db 66 VEPTSGNTGIALAYAAARGYKLTMTPEMTSTERRKLLKALGANLVLTGA-----KGM 120  
QY 140 QRTMQETGILVHPNQEPAVI-----ACQGTIALEVLNQVP-LVDALVVPVGGGG 188  
Db 121 KGAIOKAEETVADPOKYLILLOQFSXPANPEIHEKTTGPEIWDTDGQVDVFTSGVGTG 180  
QY 189 MVAGIATTKALKPSVKV--YAAEPSNADDCYQSKLKGELTPNHLHPETIADG-VKSSIG 245  
Db 181 TLTGVTTRYIKGTGKTDLTITVAVEPTSPVIAQALAGEEIKPGPHIQIGLGFIPGNLD 240

QY 246 LNTWPIIRDLVDDVFTVTDEIKYATQLVWGRMKLLIEPTAGVALAVALSQHFOQTVSPEV 305  
Db 241 LK-----LIDKVVGITNEEAISTARRLMEEGILAGISSGAVALAALKLO-EDSESFTN 292  
QY 306 KNCVIVL 312  
Db 293 KNIVWIL 299

## RESULT 10

US-09-931-457A-31  
; Sequence 31, Application US/09931457A  
; Patent No. US20020157132A1  
; GENERAL INFORMATION:  
; APPLICANT: Falco, S. Carl.  
; APPLICANT: Allen, Stephen M.  
; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes  
; FILE REFERENCE: BB1116 US CIP  
; CURRENT APPLICATION NUMBER: US/09/931,457A  
; CURRENT FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: 09/424,976  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: 60/065,385  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/049,406  
; PRIOR FILING DATE: 1997-06-12  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 31  
; LENGTH: 325  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-09-931-457A-31

Query Match 9.2%; Score 160.5; DB 9; Length 325;

Best Local Similarity 25.3%; Pred. No. 7.8e-07;  
Matches 83; Conservative 54; Mismatches 131; Indels 57; Gaps 16;

QY 12 VEKAHI--NIQDSIHLPVLTSSILNOIAG---RNLFKCELFQKTSFKIRGALNAI-- 64  
Db 3 VERSGIADKDVTELGKTPLV---YLNKLADGCVARVAAKLELMPCSSVKDRIGYSMIAD 59  
QY 65 ---RGLIPDTPEEKRAVYTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIOAY 121  
Db 60 AEEKGLI--TPGK--SVLIEPTSGNTGIGLAFMAAARGYKLIITMPASMSLERRILLAF 115  
QY 122 GASIVYCDPSDESREKVTQRMQETEGILVH-PNQEPAVIAGQ-----GTIALE 169  
Db 116 GAELVLTDP-----KMGAVQKABEILAKTN---AYILQOFENPANKVHYETTPGE 167  
QY 170 V-LNQVPLVDALVVPVGGGMVAGIATITKALKPSVKVYAAEPSNADDCYQSKLKGELTP 228  
Db 168 INKSGDKIDAFVSGIGTGTGTGAGKYLKEQNPKNKILIGVP-----VESPVLSGGKP 221  
QY 229 NLHPPETIADG-VKSSIGLNTWPIIRDLVDDVFTVTDEIKYATQLVWGRMKLLIEPTAG 287  
Db 222 GPHKIQIGAGFIPGVLEVN-----LLDEVVQISSDEAIEATAKLALKEGLFVGISG 274  
QY 288 VALAVALSQHFOQTVS-PEVKNVCIV 311  
Db 275 AAAAAA-----FQIAKRPENAGKLIV 295

## RESULT 11

US-09-931-457A-66  
; Sequence 66, Application US/09931457A  
; Patent No. US20020157132A1  
; GENERAL INFORMATION:  
; APPLICANT: Falco, S. Carl.  
; APPLICANT: Allen, Stephen M.  
; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes  
; FILE REFERENCE: BB1116 US CIP  
; CURRENT APPLICATION NUMBER: US/09/931,457A



QY 129 DPDSERKVTQ-RIMQETEGILVHPNQ-----EPAVIAGQGTIALEVLNQV--PLVDAL 180  
Db 122 PGSEAMKGAIKKAKELKEEHGYF-EPQQFENPANPEV--HELTGPELLOQFEGKTI 178  
QY 181 VVPVGGGMVAGTAITIKALKPSVKVYAAEPSNADDCYOSKLGKELTPNLHPETIADGV 240  
Db 179 LAGVGTGTLGSGVGVKVKKEYPNIEVAIEPE-----ASPVLSGGEGPGPKLOGLGAGF 232  
QY 241 KSSIGLNTWPIIRDLDVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQ 296  
Db 233 IPGT-LNT-----EIVDSIIKVGNDTAMEMSRRAKEEGILAGISSGAAIYAAIQ 282

## RESULT 14

US-09-815-242-12800  
; Sequence 12800, Application US/09815242  
; Patent No. US20020061569A1

## GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 12800  
; LENGTH: 313  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus

US-09-815-242-12800

Query Match 9.2%; Score 159.5; DB 10; Length 313;  
Best Local Similarity 23.8%; Pred. No. 9.1e-07;  
Matches 70; Conservative 52; Mismatches 151; Indels 23; Gaps 8;  
QY 9 FADVEKAHINQDSIHLPVLTSSILNQIAGRNLFKFCELFQKTSFKIRGALNAIRGLI 68  
Db 2 FLMAQRPVDNITQIGTTPVKLRNVDDNADVYVKLEYQNPQGSVKDRIALAMIEKAE 61  
QY 69 PDTPEKPKAVYTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIOAYGASIVYC 128  
Db 62 REGKIRPGDITVEPTSGNTGIGLAFYCAAKGYKAVFTMPETMSQERRNLLKAYGAELVLT 121  
QY 129 DPDSERKVTQ-RIMQETEGILVHPNQ-----EPAVIAGQGTIALEVLNQV--PLVDAL 180  
Db 122 PGSEAMKGAIKKAKELKEEHGYF-EPQQFENPANPEV--HELTGPELLOQFEGKTI 178  
QY 181 VVPVGGGMVAGTAITIKALKPSVKVYAAEPSNADDCYOSKLGKELTPNLHPETIADGV 240  
Db 179 LAGVGTGTLGSGVGVKVKKEYPNIEVAIEPE-----ASPVLSGGEGPGPKLOGLGAGF 232

QY 241 KSSIGLNTWPIIRDLDVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQ 296  
Db 233 IPGT-LNT-----EIVDSIIKVGNDTAMEMSRRAKEEGILAGISSGAAIYAAIQ 282

## RESULT 15

US-09-815-242-5437  
; Sequence 5437, Application US/09815242  
; Patent No. US20020061569A1

## GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 5437  
; LENGTH: 303  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus

US-09-815-242-5437

Query Match 9.1%; Score 157.5; DB 10; Length 303;  
Best Local Similarity 23.7%; Pred. No. 1.3e-06;  
Matches 68; Conservative 51; Mismatches 145; Indels 23; Gaps 8;  
QY 18 NIQDSIHLPVLTSSILNQIAGRNLFKFCELFQKTSFKIRGALNAIRGLIPTPEKPK 77  
Db 2 NITQIGGTPVKLRNVDDNADVYVKLEYQNPQGSVKDRIALAMIEKAEKIRPGD 61  
QY 78 AVYTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIOAYGASIVYCDPSDESREK 137  
Db 62 TIVEPTSGNTGIGLAFYCAAKGYKAVFTMPETMSQERRNLLKAYGAELVLTGSEAMKGA 121  
QY 138 VTQ-RIMQETEGILVHPNQ-----EPAVIAGQGTIALEVLNQV--PLVDALVVPVGGGM 189  
Db 122 IKKAKELKEEHGYF-EPQQFENPANPEV--HELTGPELLOQFEGKTI 178  
QY 190 VAGTATITKALKPSVKVYAAEPSNADDCYOSKLGKELTPNLHPETIADGVKSSIGLNTW 249  
Db 179 LSGVGKVLKKEYPNIEVAIEPE-----ASPVLSGGEGPGPKLOGLGAGFPGT-LNT- 230  
QY 250 PIIRDLDVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQ 296  
Db 231 ----EIVDSIIKVGNDTAMEMSRRAKEEGILAGISSGAAIYAAIQ 273

Search completed: June 24, 2003, 06:16:49  
Job time : 277.093 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 06:07:29 ; Search time 26.9602 Seconds  
(without alignments)  
1208.801 Million cell updates/sec

Title: US-09-889-609B-8  
Perfect score: 1740  
Sequence: 1 MCAQYCISPADVEKAHINIQ.....TSLNWGQAERPAPQTQVSV 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	791	45.5	346	2 T04211	probable threonine
2	587	33.7	326	2 S38061	threonine ammonia-
3	560	32.2	323	2 T41297	threonine ammonia-
4	544	31.3	320	2 E83310	L-serine ammonia-1
5	535	30.7	346	2 F89921	threonine ammonia-
6	530	30.5	325	2 T36434	threonine ammonia-
7	522.5	30.0	401	2 D72386	threonine ammonia-
8	520.5	29.9	324	2 A97587	probable serine/th
9	520.5	29.9	324	2 AF2312	threonine dehydrat
10	517.5	29.7	323	2 F95390	probable threonine
11	507.5	29.2	317	2 T24337	threonine ammonia-
12	499	28.7	405	2 D90166	hypothetical prote
13	498.5	28.6	325	2 G87633	threonine dehydrat
14	492.5	28.3	495	2 H84359	threonine dehydrat
15	491	28.2	329	2 AB0897	threonine ammonia-
16	476	27.4	329	1 DWECDT	threonine ammonia-
17	476	27.4	329	2 D85973	threonine ammonia-
18	476	27.4	329	2 E91128	threonine ammonia-
19	466.5	26.8	320	2 AG0496	threonine ammonia-
20	466.5	26.8	343	2 D97778	threonine ammonia-
21	463.5	26.6	504	2 F83603	threonine dehydrat
22	446.5	25.7	403	2 D81355	threonine ammonia-
23	444.5	25.5	333	2 H71703	probable threonine
24	441.5	25.4	568	2 E75502	threonine ammonia-
25	439	25.2	400	2 AB7700	threonine dehydrat
26	426.5	24.5	336	2 AC3142	threonine dehydrat
27	423	24.3	512	2 T23166	hypothetical prote
28	414	23.8	334	2 B95894	probable threonine
29	413	23.7	600	2 T39516	threonine ammonia-

30	409.5	23.5	508	2 S77559	threonine ammonia-
31	406.5	23.4	515	2 B83479	threonine dehydrat
32	403.5	23.2	304	2 A98146	probable threonine
33	399	22.9	349	2 C72630	threonine ammonia-
34	398	22.9	514	2 AG0474	threonine ammonia-
35	395.5	22.7	576	1 DWBYT	threonine ammonia-
36	391	22.5	510	2 E82374	threonine ammonia-
37	384	22.1	514	2 AB0924	threonine ammonia-
38	382	22.0	514	2 B91217	threonine ammonia-
39	380	21.8	514	1 DWEETT	threonine ammonia-
40	378.5	21.8	515	2 C86063	threonine ammonia-
41	378	21.7	514	1 DWEETS	threonine ammonia-
42	374	21.5	422	2 A69644	threonine ammonia-
43	371.5	21.4	595	2 A38628	threonine ammonia-
44	371	21.3	592	2 T51712	threonine ammonia-
45	368	21.1	257	2 AF1839	threonine dehydrat

ALIGNMENTS

RESULT 1

T04211  
Probable threonine ammonia-lyase (EC 4.3.1.19) T5C23.70 [similarity] - Arabidopsis th  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 21-Jun-2002  
C:Accession: T04211  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z15261  
A:Accession: T04211  
A:Molecule type: DNA  
A:Residues: 1-346 <BEV>  
A:Cross-references: EMBL:AL049500  
A:Experimental source: cultivar Columbia; BAC clone T5C23  
C:Genetics:  
A:Map position: 4  
A:Introns: 55/2; 99/2; 176/2; 263/2  
A:Note: T5C23.70  
C:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase

Query Match 45.5%; Score 791; DB 2; Length 346;  
Best Local Similarity 46.9%; Pred. No. 2.3e-54;  
Matches 159; Conservative 68; Mismatches 88; Indels 24; Gaps 6;

Qy	4	QYCISFADVEKAHINIQDSIHLTPVLTSSILNQIAGRNLFKCELFQ-----	50
Db	7	KYAADILSIKEAHDRIKPYIHRTPVLTSELSNLSGRSLFFKCECLQKGIERSFYESVY	66
Qy	51	--KTGSFKIRGALNIRGLIPDTPPEKPAVVTSSNHNQALTYAAKLGIPAYIVVPO	108
Db	67	CGKSGAFKFGAGCNVLSL---DAEQAAKGVTSSNHNQALTYAAKLGIPAYIVVPO	123
Qy	109	TAPNCKKLAIQAYCASIVYCDPDESREKVTQIMQETEGILVHPNQPAPVIAQOGTIAL	168
Db	124	GAPCKVDNVIRYGGKVINSEATMSREETASKVLQETGSLVHPYNDGRIISQGTIAL	183
Qy	169	EVLNQVPLVDALVVPVGGGGMVAGIATIKALPKSVKVAAPSNADDCYQSKLKGELTP	228
Db	184	ELLEQIEIDAIIVVPISSGGLISGVALAAKSIKPSIRITAAEPKGAADAAQSKVAGII-	242
Qy	229	NLHPPETIADGVKSSIGLNTWPIIRDLVDVFTVEDEIKYATQLVWGRMKLLIEPTAGV	288
Db	243	TLPTVTIADGLRASLDLTPVVPVRLVDVDTVEECEIIEAMKMCYIILKVSVEPSGAI	302
Qy	289	ALAAVLSQHFQTVSP---EVKNVICVLSGGNVDLTSLNW	324
Db	303	GLAAVLSNFRN-NPSCRCKNIGVLSGGNVDLGS-LW	339

RESULT 2

S38061

threonine ammonia-lyase (EC 4.3.1.19) YKL218c [similarity] - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein D326  
C:Species: Saccharomyces cerevisiae  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 21-Jun-2002.  
C:Accession: S38061; S44320  
R:Alexandraki, D.; Tzeremia, M.  
submitted to the Protein Sequence Database, March 1994  
A:Reference number: S38054  
A:Accession: S38061  
A:Molecule type: DNA  
A:Residues: 1-326 <ALE>  
A:Cross-references: EMBL:Z28218; MID:g486390; PIDN:CAA82063.1; PID:g486391; MIPS:YKL218c  
A:Experimental source: strain S288C  
R:Tzeremia, M.; Horaitis, O.; Alexandraki, D.  
Yeast 10, 663-679, 1994  
A:Title: The complete sequencing of a 24.6 kb segment of yeast chromosome XI identified  
e dehydratases, membrane transporters, hydantoins and the phospholipase A(2)-activat  
A:Reference number: S44319; MUID:95028164; PMID:7941750  
A:Accession: S44320  
A:Molecule type: DNA  
A:Residues: 1-326 <T2E>  
A:Cross-references: EMBL:X75951; MID:g473130; PIDN:CAA53555.1; PID:g473132  
C:Genetics:  
A:Gene: SGB:SRY1  
A:Cross-references: SGD:S0001701  
A:Map position: 11L  
C:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase  
  
Query Match 33.7%; Score 587; DB 2; Length 326;  
Best Local Similarity 40.9%; Pred. No. 2e-38;  
Matches 128; Conservative 66; Mismatches 111; Indels 8; Gaps 4;  
  
QY 8 SFADVEKAHINQDSIHLTPVLTSSILNQIAGRNLFKFCFELFQKTSFKIRGALNAIRGL 67  
DB 5 TYGDVLDASNRKEYNKTPVLTSSILNQIAGRNLFKFCFELFQKTSFKIRGALNAIRGL 64  
QY 68 IPDTPEEKPAKVYTHSSGNHQAALTYAAKLEGIPAVIVPQTPAPNCKKLAIAQYASIVY 127  
DB 65 ---SDEKRSKGVIAFSSGNHQAALTYAAKLEGIPAVIVPQTPAPNCKKLAIAQYASIVY 121  
QY 128 CDPDSRESKVTORIMQETEGILVHPNQEPVAVTAGGTTALEVLNQVPLVDALVVPVGG 197  
DB 122 YNYRTEDREGIQRLAAEHGFAIPYDHPDVIAGGTSKELLEVEVGQDLDALFVLGG 181  
QY 188 GWAGTAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVKSS-IGL 246  
DB 182 GLSSGALARSLSPECKIFGVPEAGNDCQSQFRGSIV-HINTPKTIADGQTOHLGE 240  
QY 247 NTWPIIRDLVDVFTTEDEIKYATQLVGMKMLLEPTAGVALAVALSQHFTVSPEVK 306  
DB 241 YTEAITIRENVDDILTYSQDELVKCMHFLAERMKVVVEPTACLGAFAGALLKKEELVG--K 297  
QY 307 NVCIVLSGGNVDL 319  
DB 298 KVGIIISGGNVDM 310  
  
RESULT 3  
T41297  
Threonine ammonia-lyase (EC 4.3.1.19) SPCC320.14 [similarity] - fission yeast (Schizosac  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jun-2002  
C:Accession: T41297; T41325  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Wedler, H.; Wambutt, R.; Wedler, E.  
submitted to the EMBL Data Library, March 1998  
A:Reference number: Z21986  
A:Accession: T41297  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-323 <WOO>  
A:Cross-references: EMBL:AL022245; PIDN:CAAL8316.1; GSPDB:GN00068; SPDB:SPCC320.14  
A:Experimental source: strain 972h-; cosmid c320

R;Gwilliam, R.; Barrell, B.G.; Rajandream, M.A.; Wedler, H.; Wambutt, R.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z21987  
A:Accession: T41325  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-323 <GWI>  
A:Cross-references: EMBL:AL031603; PIDN:CAA20920.1; GSPDB:GN00068; SPDB:SPCC330.15c  
A:Experimental source: strain 972h-; cosmid c330  
C:Genetics:  
A:Gene: SPCC320.14; SPCC330.15c  
A:Map position: 3  
C:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase  
  
Query Match 32.2%; Score 560; DB 2; Length 323;  
Best Local Similarity 39.6%; Pred. No. 2.6e-36;  
Matches 125; Conservative 68; Mismatches 109; Indels 14; Gaps 7;  
  
QY 8 SFADVEKAHINQDSIHLTPVLTSSILNQIAGRNLFKFCFELFQKTSFKIRGALNAIRGL 67  
DB 9 TYDDVASASERIKKFKANKTPVLTSSIVNKEFVAEVEFKENFOMGAFAKFRGALNALSQ 68  
QY 68 IPDTPEEKPAKVYTHSSGNHQAALTYAAKLEGIPAVIVPQTPAPNCKKLAIAQYASIVY 127  
DB 69 ---NEAQRKAGVLTSSGNHQAALTYAAKLEGIPAVIVPQTPAPNCKKLAIAQYASIVY 125  
QY 128 CDPDSRESKVTORIMQETEGILVHPNQEPVAVTAGGTTALEVLNQVPLVDALVVPVGG 186  
DB 126 YDRYKDDREKMAKEI-SEREGLTIIPIYDHPVHVLGAGGTAAKELFEVEGPDALFVCLGG 184  
QY 187 GGVAGTAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVKSS-IG 245  
DB 185 GGLSSGALARSLSPECKIFGVPEAGNDCQSQFRGSIV-HIDTPKTIADGQTOHLG 243  
QY 246 LNTWPIIRDLVDVFTTEDEIKYATQLVGMKMLLEPTAGVALAVALSQHFTVSPEVK 305  
DB 244 NYTFSITIKKVDILTYSQDELVKCMHFLAERMKVVVEPTGCLSFAAA-----RAMKEK 298  
QY 306 KN--VCIVLSGGNVDL 319  
DB 299 KNKRIGIIISGGNVDI 314  
  
RESULT 4  
E83310  
L-serine ammonia-lyase (EC 4.3.1.17) [similarity] - Pseudomonas aeruginosa (strain PA  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 21-Jun-2002  
C:Accession: E83310  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: E83310  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-320 <STO>  
A:Cross-references: GB:A004696; GB:A004091; MID:g9948750; PIDN:AAG06071.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2683  
C:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase  
  
Query Match 31.3%; Score 544; DB 2; Length 320;  
Best Local Similarity 39.0%; Pred. No. 4.7e-35;  
Matches 122; Conservative 62; Mismatches 121; Indels 8; Gaps 4;  
  
QY 8 STADVEKAHINQDSIHLTPVLTSSILNQIAGRNLFKFCFELFQKTSFKIRGALNAIRGL 67  
DB 9 TYDDVASASERIKKFKANKTPVLTSSIVNKEFVAEVEFKENFOMGAFAKFRGALNALSQ 68

Db 6 TYDDVIAAARIAIAGHANRTVMSSRTLDEELGAEVFFKCNELQRMGMFAKFRGAFNALSRF 65  
QY 68 IPDTPPEKPAVTVTHSSGNHGOALTAAKLEGIPAYIVVQTPAPNCKKLAIAQYASIV 127  
Db 66 ---SAGORAGVAVFSSGNHQAIALSARLLGIPATIVMPADAPAVKIENATRYGGQVVL 122  
QY 128 CDPDSRESKVTQRIQWETEGILVHPNQEPAVIAGOGTTALEVLNVLVPLDALVVPVGG 187  
Db 123 YDRTYEDREQIGDLAQRLHGLTLPYDHPDVLVLAGOGTAAKELFEVGLDFAFFLPGG 182  
QY 188 GWVAGTAITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPETADGVKSS-IGL 246  
Db 183 GLUSGALAIARAPACRIYGVPEAGNDGQSRSGAIV-HDITPTTADGATQHLGN 241  
QY 247 NTWPIIRDLVDDVFTTEDEIKYATQLVWGRMKLLIETPAGVALAAVLSOHFOTVSPEVK 306  
Db 242 LTFPLIQRNVDDILTASDAELVDGMRFLAARMKLLVEPTGCLGLAAARQKDEL---RGK 298  
QY 307 NVCIVLSGGNVDL 319  
Db 299 RVGILLSGGNVDL 311

RESULT 5  
F89921  
Query Match 30.7%; Score 535; DB 2; Length 346;  
Best Local Similarity 37.8%; Pred. No. 2.6e-34;  
Matches 122; Conservative 64; Mismatches 131; Indels 6; Gaps 4;  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 02-Aug-2002  
C:Accession: F89921  
R:Kuroda, M.; Ohca, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud  
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: F89921  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-346 <KUR>  
A:Cross-references: GB:BA000018; PID:g13701236; PIDN:BABA2531.1; GSPDB:GNO0149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA1271  
A:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase

QY 5 YCISFADVEKAHINIQDSTHLTPVLTSSILNQ-IAGRNLFKCELFQKTSFKIRGALNA 63  
Db 12 HIVSLGDIIEAKASIKPFIRTPPLIKSWLSQNTKGNVYKLENMQFTGSGFRGASNK 71  
QY 64 IRLGIPDTPPEKPAVTVTHSSGNHGOALTAAKLEGIPAYIVVQTPAPNCKKLAIAQYGA 123  
Db 72 INHL---SDQKAKGIIGASAGNHAQGVALTAKLLGIDATIVMPETAPIAKONATKGYGA-128  
QY 124 SIYCDPSDESREKVTQRIQWETEGILVHPNQEPAVIAGOGTTALEVLNVLVPLDALVVP 183  
Db 129 KVLKGNFNETRLYMEELAKENGMTIVHPYDKFVMAGOGTIGLEILDDIMNVNVIIVP 188  
QY 184 VGGGMVAGTAITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPETADGVKSS 243  
Db 189 VGGGLIAGIATALKSFNSIHIIGVQAEVNHGMAESFYKRALTEH-REDSTIADGCDVK 247  
QY 244 I-GLNTWPIIRDLVDDVFTTEDEIKYATQLVWGRMKLLIETPAGVALAAVLSOHFQTVS 302  
Db 248 VPCEKTYEVKHLVDFEILVSEIEHAMQDLQRAKIITEGAGALPTAAILSGKIDKKW 307  
QY 303 PEVKNVCIVLSGGNVDLTSLNWV 325  
Db 308 LEGKNVALVSGGNVDLTRVSGV 330

RESULT 6  
T36434  
Query Match 30.5%; Score 530; DB 2; Length 325;  
Best Local Similarity 39.3%; Pred. No. 6e-34;  
Matches 123; Conservative 54; Mismatches 128; Indels 8; Gaps 4;  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jun-2002  
C:Accession: T36434  
R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, July 1999  
A:Reference number: Z21598  
A:Accession: T36434  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-325 <SEE>  
A:Cross-references: EMBL:AL096837; PIDN:CABA48898.1; GSPDB:GNO0070; SCOEDB:SCF43A.11c  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCF43A.11c  
C:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase

QY 7 ISFADVEKAHINIQDSTHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGALNAIRG 66  
Db 8 VTLDVRSAAARIRGVAHRTPTLRSTRLDALVGAEVHLKCNQORVGAFFRGAAYNAASR 67  
QY 67 LIPDTPPEKPAVTVTHSSGNHGOALTAAKLEGIPAYIVVQTPAPNCKKLAIAQYASIV 126  
Db 68 L---TPEQLARGIAAYSSGNHQAQVAALARELGTAVIVMPEDAPPSKRDATAGYGAIV 124  
QY 127 YCDPSDESREKVTQRIQWETEGILVHPNQEPAVIAGOGTTALEVLNVLVPLDALVVPVGG 186  
Db 125 TYDRYTGDRVAVAAALAAADRLTLIPPYEHPPHYIAGOGTAALVELVEETGELDALVAPVG 184  
QY 187 GGMVAGTAITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPETADG-VKSSIG 245  
Db 185 GGLIAGSATAVKALHFQMGVIGVEPEAGDDTRSLANGRRV-SVPVPTIADQALPTPG 243  
QY 246 LNTWPIIRDLVDDVFTTEDEIKYATQLVWGRMKLLIETPAGVALAAVLSOHFQTVSPEV 305  
Db 244 ELTFSLNRRLDGLVLDSDDEIRDAIRFAFERLKTVLEPSGATPLAALLNGRIDALP--- 300

RESULT 7  
D72386  
Query Match 30.6%; Score 530; DB 2; Length 325;  
Best Local Similarity 39.3%; Pred. No. 6e-34;  
Matches 123; Conservative 54; Mismatches 128; Indels 8; Gaps 4;  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jun-2002  
C:Accession: D72386  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garrett, M.E.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: D72386  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-401 <ARN>  
A:Cross-references: GB:AE001716; GB:AE000512; NID:g4980853; PIDN:AAD35443.1; PID:g498  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TW0356  
C:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase; p

```

F:51/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match      30.0%  Score 522.5;  DB 2;  Length 401;
Best Local Similarity 35.8%;  Pred. No. 3.1e-33;
Matches 114;  Conservative 72;  Mismatches 123;  Indels 9;  Gaps 4;

Qy 7 ISADVKEAHINIQDSTIHLPVLTSSILNQIAGRNLFFKCELFQKTSFKIRGALNAIRG 66
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2 ITLEDKEAORTLKNVVRHTALAYSSVLSVTGGEIYLNKMNLEQKTSFKIRGAYNKIAH 61
Qy 67 LIPDTPKEKFAVYTHSSGNHGQALTYAAKLEGIPAVIVVPQTPANCKKLAICAYGASIV 126
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 L---SEERKRGVVAASAGNHAQVLAAGIFGIPATIVMPRYAPLSKTKTKTNLGAQVI 118
Qy 127 YCDPSDESREKVTQRIMQETEGILVHPNQEPAVIAGQGTIALEVLNQPLVDALVVPVG 186
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
119 LEGNIFDEAYEALRIQETGAVFVHPFNPHVIAQGGTIGLEIMEDLPDVEVVVPVG 178
Qy 187 GGVVAGTATITKALKFSVKYAAEPSNADDCCYOSKLGELTPNLHPETIADGVK--SSIG 245
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
179 GGLISGVSAVKAISMNPEVKVIGYQTNMPSMIASLRGR--AERVEGKPTLADGIAVKKPG 237
Qy 246 LNTWPIIRDLVDVFTTDEIEKYATQLVNMRKLLITEPTAGVALAAILSOHQFTVSPEV 305
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
238 DLFLVKKYVDENVAVNEEIEADILFLLEQAKVVAEGAGAVGVAVLNK-----LDVK 293
Qy 306 KNCVILSGNGVDLTSLN 323
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
294 KKVAISVGGNIDWNMID 311
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 8  
A97687  
probable serine/threonine dehydratase, degradative (PA2683) [imported] - Agrobacterium t  
C:Species: Agrobacterium tumefaciens  
C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: A97687  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurolo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: A97687  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-324 <RUR>  
A:Cross-references: GB:AE007869; PIDN:AAK88450.1; PID:gl5157951; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C\_4956  
A:Map position: circular chromosome  
C:Superfamily: threonine dehydratase

Query Match	29.9%	Score	520.5	DB 2	Length	324			
Best Local Similarity	40.6%	Pred. No.	3.3e-33						
Matches	127	Conservative	44	Mismatches	129	Indels	13	Gaps	5
QY	12	VEKAHINIODSIHLTPVLTSSILNQTAGRNLFKFCLEFQKTSFKIRGALNATRGILPOT	71						
Db	7	IEAAREIGNHAYRTPPLLTSPFIDEIAGRKLFKVAECLOQTGSKFRGWSAVSGILPADV	66						
QY	72	PEKPRAVVTHSSGNHGQALTYAAKLEGIPAYIVVPOTAPNCKKLAIQAYASIVYCDPS	131						
Db	67	---RAKGVIATFSSGNAHQGVALAARLHGIPAVIIMPSPDKIKIDNTRAYGAEEVLYDRA	123						
QY	132	DESREKVTORIMOTEGILVHPHNOEPAVTAGQCTIALEVLNQ----	186						
Db	124	NEDRDAIGNRLSSERGLTIRPYDEPLVYTAGQTAGLTAEQGAELGIGAAEVL-VPCCG	182						
QY	187	GMVAGIATITKALKPSVKVYAAEPSNADDCYOSKUGELTPTNLHPPTIADG-VKSSGT	245						
Db	183	GLGTSGISALDAAKARNYKVRTAEPRFFDDVARSAAAGKIERNATTSGSGICDAIVTPQPG	242						
QY	246	LNTWPRIIRLDVDFVTVTDEIKYATQLVWGRMKLLIETPTAGVAAALAVLSQHFOFVSPW	305						

```

      :|::| | :||| | :|::||| ||||| |
Db    243 NITFPIMAGCGXGIAVSEEARVLAFNRKLVIEPGAVAAAL---FHGRELES 299

Qy    306 KNVCIVLGGNVD 318
      : | | |||||
Db    300 ETVIYASGGNVD 312


RESULT 9'
AF2912
threonine dehydratase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AF2912
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woode-
rage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCI
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam-
ster, E.W.
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AF2912
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-324 <Kur>
A:Cross-references: GB:AE008688; PID:NAL43716.1; PID:g17741247; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
```

```

Query Match          29.9%; Score 520.5; DB 2; Length 324;
Best Local Similarity 40.6%; Pred. No. 3.3e-33;
Matches 127; Conservative 44; Mismatches 129; Indels 13; Gaps 5;

Qy      12 VEKAHINIODSIHLTPVLTSIIINQIAGRNLFKFCLEFQKTGSKIRGALNAIRGLIPOT 71
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      7 IEAARERIGNHAVRTPLLTSPFLDIAGRKLFVKAECQLQTGSKFKPRGWSAVSGLPADV 66

Qy      72 PEKPKRAVTHSSGNHGQALTYAAKLEGIPAYIVVPOTAPNCKKLAIQAYGASIVYCDPS 131
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      67 ---RAKGVIASFSGNHAQGVAAARLHGIPAVIIMPSPADPKIKIDNTRAYGAEEVLYDRA 123

Qy      132 DESREKVTQIRIMOETGECILVHPNOEPAVTAGOQTIALEVLNQ-----VPLVDALVVPVGG 186
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      124 NEDRDIAIGNLSSLSERGTLIRPYDEPLVIATAGQQTAGLETAEOGAELGIGAAEVL-VPCGG 182

Qy      187 GGMVAGIATIKALKPSVKVYAAEPSNADDCYSKLGELTPNLHPPETIADG-VKSSSTG 245
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      183 GGLTSGISLADAKARNYKVRTAEPRFDVVARSLAAGRIERNATTSSGSDCAIVTPQPG 242

Qy      246 LNTWPIIRDLVDVDFVTWTEDEIKYATQLVWGRMKLLIETAGVALAASLSQHIFQTVSPVE 305
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      243 NIITFPIMAGLCGGIASEEALRAMVLAFNRLLKVVIEPCGAVALAAL---FHGKELES 299

Qy      306 KNCIVLSGGNVD 318
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      300 ETVIAVAGSNVD 312
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 10
F95390
C:Species: Sinorhizobium meliloti [imported] - Sinorhizobium meliloti (strain 1021) maga
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 27-Nov-2001
C:Accession: F95390
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; B
; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Suryeckti, R.; Wells, D.H.; Yeh, K
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9886, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli
A:Reference number: A95262 MUID:21396509; PMID:11481432

```

A:Accession: F95390  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-323 <KUR>  
A:Cross-references: GB:AE006469; PIDN:AAK65688.1; PID:g14524179; GSPDB:GN00165  
A:Experimental source: strain 1021, megaplasmid pSYMA  
R:Galibert, F.; Finan, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
pela, D.; Chain, P.; Finan, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: Sma1872  
A:Genome: plasmid  
C:Superfamily: threonine dehydratase

Query Match 29.7%; Score 517.5; DB 2; Length 323;  
Best Local Similarity 37.7%; Pred. No. 5.6e-33;  
Matches 118; Conservative 60; Mismatches 128; Indels 7; Gaps 4;

QY 11 DVEKAHINQDSIHLPVTSSILNQIAGRNLFKCELFQKTSFKIRGALNAIRGLIPD 70  
DB 9 DIREARELRKPHVHTPLRAEKIEKAGACQLYKPTLQITGAFKIRGALKALS--- 65

QY 71 TPEKPKAVVTHSSNGHGOALTYAAKLEGIPAYIVVTPQAPNCKKLAIOAYGASIVYCDP 130  
DB 66 SREEIANGIATSSNGHAGLSYAAKMLGVKVLVLPVTPPKIKIENTKALGAEVILFDG 125

QY 131 SDESREKVTORIMQETEGILVHPNQEPVIAAGQTTALEVLNQVPLVDALVVPVGGGGMV 190  
DB 126 DNAARKKKVEAEGNKYAVHGFEDPVVMAGOGTIGCELEDLDVDVIVPLGGGLI 185

QY 191 AGIATIKALKSVKYAAEPSNADDCYOSKLGELTPNLHPETIADGVKSSI-GLNTW 249  
DB 186 SGATAIKETKPSVRVIGAEPAITPKYFHSRVNKERT-SLPLKNTIADGLRISVPGQNPY 244

QY 250 PIIRDLVDVFWTEDEIKYATQLVWGRMKLLIEPTAGVALAVALSOHFQTVSPEKVCV 309  
DB 245 PIIEKVVDEIVLVEDHIIAGHRAKADKLAIEPAASIGVALLAGIIVDKLDE--KVC 302

QY 310 IVLSSGNVDLTSL 322  
DB 303 AVLSSGNWDLRLD 315

RESULT 11  
T24337  
threonine ammonia-lyase (EC 4.3.1.19) T01H8.2 [similarity] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jun-2002  
C:Accession: T24337  
R:Lennard, N.  
submitted to the EMBL Data Library, September 1996  
A:Reference number: Z19877  
A:Accession: T24337  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-317 <MIL>  
A:Cross-references: EMBL:Z80219; PIDN:CAB02298.1; GSPDB:GN00019; CESP:T01H8.2  
A:Experimental source: clone T01H8  
C:Genetics:  
A:Gene: CESP-T01H8.2  
A:Map position: 1  
A:Introns: 75/3; 131/2; 159/3; 203/1; 259/3; 276/3  
C:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase; phos  
F:56/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 29.2%; Score 507.5; DB 2; Length 317;

Best Local Similarity 37.7%; Pred. No. 3.3e-32;  
Matches 121; Conservative 62; Mismatches 115; Indels 23; Gaps 9;

QY 7 ISFADVEKAHINQDSIHLPVTSSILNQIAGRNLFKCELFQKTSFKIRGALNAI 64  
DB 5 ITWENRAAYQRTAPFLHKTIVITSENIDEKGVNGTHVLFKCEHLOKTSFKARGALNS- 63

QY 65 RGLIPDTPEEKPAVVTHSSNGHGOALTYAAKLEGIPAYIVVTPQAPNCKKLAIOAYGAS 124  
DB 64 -AIL--AKEKNAGMIAHSSNGHGOALAWAAQKIGLPTIVVKNAPISKIEGMREYN 120

QY 125 IYVCDPSDESREKVTORIMQETEGILVHPNQEPVIAAGQTTALEVLNQV-PLVDALVVP 183  
DB 121 IVFCEPTVTSRESVCADLTETKLEYCYEPCVNCVMINGSSVAFELIQQVGNEDISIFLS 180

QY 184 VGGGMVAGIATITKALKPSVKVYAAEPSNADDCYOSKLGELTPNLHPD---ETIADG 239  
DB 181 VGGGLASSVAFELIGNLRPDIEVLYVQA-----QKELSNFLENGIKCPVDFLTIADG 234

QY 240 VK-SSIGLNTWPIIRD-LVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAVALSOH 297  
DB 235 VRVARVGPLCEPILKQYCAQKVISVKEDEITKEGLKLIWTRMKQRIEPTAALAFAGVLYHK 294

QY 298 FQTVSPEVKNVCIIVSSGGNVD 318  
DB 295 ----PAHIRPLVILCGGNVD 311

RESULT 12  
D90166  
hypotheical protein tdcB [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Jun-2002  
C:Accession: D90166  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder  
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: D90166  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-405 <KUR>  
A:Cross-references: GB:AE006641; NID:g13813385; PIDN:AAK40587.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: tdcB  
C:Superfamily: threonine dehydratase

Query Match 28.7%; Score 499; DB 2; Length 405;  
Best Local Similarity 39.1%; Pred. No. 2.2e-31;  
Matches 126; Conservative 60; Mismatches 124; Indels 12; Gaps 8;

QY 5 YCISFADVEKAHINQDSIHLPVTSSILNQIAGRNLFKCELFQKTSFKIRGALNAI 64  
DB 3 YLEFDRIRLAKEKIEKYVHITPIDYSTFTFSIRINAKVYLKLENLOKTSFKYRGAFNKL 62

QY 65 RGLIPDTPEEKPAVVTHSSNGHGOALTYAAKLEGIPAYIVVTPQAPNCKKLAIOAYGAS 124  
DB 63 LSL---KEEKNKNCVIAVSAGNHAQGVAYAASTLNKSTIVMPETAPASKYLATKSYGAE 119

QY 125 IV-YCDPSDESREKVTORIMQETEGILVHPNQEPVIAAGQTTALEVLNQVPLVDALVVP 183  
DB 120 VVLGYKLYHESMKK-AEELIQNTGLIFVHPYSLDVITGGTIGLELYDIEP--DVYIIP 176

QY 184 VGGGMVAGIATITKALKPSVKVYAAEPSNADDCYOSKLGELTPNLHPETIADG--VK 241  
DB 177 IGGGLGISGIALKYRFPFNKIIIGVQSSSPSMKYSKDLGRV-EIEPSYSTADILVK 235

QY 242 SSIGLNTWPIIRDVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAVALSOHFQTV 301  
DB 236 SPSEL-TFSIINELVDDIVLVDDEETAEALVLLERSKTLAEGAGAAALASLISGVK-V 293



```

||||:||||| ||: ||: | : | ||: | ||: | : ||| |
Db 184 GGGLIAGIAIAIKSINPTIKVIGOAENVHGMAASYAGEITAH-RTTGTILADGCDVSR 242
QY 244 IGLNTPIIIRDLVDVFTVTEDEIKYATOLVWGRMKLLJEPTAGVALAAVLS-----Q 296
Db 243 PGNLYEIVRELVDDIVLYSEDEIRNSMLALIQRNKVITEGAGALACAALLSGKLDSHIQ 302
QY 297 HFQTVSPEVKNVCIVLSGCGNVDLTSLNWV 325
Db 303 NRKTVS-----IISGNIIDLSRVSQI 323
```

Search completed: June 24, 2003, 06:23:09  
Job time : 28.9602 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:44:58 ; Search time 41.4389 Seconds  
(without alignments)  
1685.614 Million cell updates/sec

Title: US-09-889-609B-8  
Perfect score: 1740  
Sequence: 1 MCAQYICISFADVEKAHINIQ.....TSLNVWGOAERPAPYQTVSV 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp-archaea.\*  
2: sp-bacteria.\*  
3: sp-fungi.\*  
4: sp-human.\*  
5: sp-invertebrate.\*  
6: sp-mammal.\*  
7: sp-mhc.\*  
8: sp-organelle.\*  
9: sp-phage.\*  
10: sp-plant.\*  
11: sp-rodent.\*  
12: sp-virus.\*  
13: sp-vertebrate.\*  
14: sp-unclassified.\*  
15: sp-rvivirus.\*  
16: sp-bacteriap.\*  
17: sp-archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	791	45.5	346	10 Q9T0D1	Q9T0d1 arabidopsis
2	620	35.6	323	16 Q8XPX0	Q8xpx0 ralstonia s
3	546	31.4	388	3 Q8X0J0	Q8x0j0 neurospora
4	544	31.3	320	16 Q910F5	Q910f5 pseudomonas
5	535	30.7	346	16 Q99U50	Q99u50 staphylococ
6	530	30.5	325	16 Q9XAA4	Q9xaa4 streptomyce
7	522.5	30.0	401	16 Q9WYJ1	Q9wyj1 thermotoga
8	520.5	29.9	324	16 Q8UBW5	Q8ubw5 agrobacteri
9	517.5	29.7	323	16 Q92Y58	Q92y58 rhizobium m
10	507.5	29.2	317	5 Q93968	Q93968 caenorhabdi
11	502.5	28.9	404	17 Q975Y3	Q975y3 sulfolobus
12	501	28.8	402	16 Q8XL77	Q8x177 clostridium
13	499	28.7	405	17 Q980P1	Q980p1 sulfolobus
14	498.5	28.6	325	16 Q9A3U7	Q9a3u7 caulobacter
15	492.5	28.3	495	17 Q9HNH6	Q9hnh6 halobacteri
16	491	28.2	329	16 Q8Z3K4	Q8z3k4 salmonella

17	487	28.0	404	16 Q8RDT9	Q8rdt9 fusobacteri
18	483.5	27.8	409	16 Q9EWG3	Q9ewg3 streptomyce
19	470	27.0	406	17 Q97CB2	Q97cb2 thermoplasm
20	468	26.9	406	17 Q9HLW2	Q9hlw2 thermoplasm
21	466.5	26.8	320	16 Q8Z9V3	Q8z9v3 yersinia pe
22	466.5	26.8	343	16 Q9ZH22	Q9zh22 rickettsia
23	464	26.7	403	17 Q8ZVF0	Q8zvf0 pyrobaculum
24	463.5	26.6	504	16 Q916G0	Q916g0 pseudomonas
25	453	26.0	469	5 Q9VHF0	Q9vhf0 drosophila
26	446.5	25.7	403	16 Q9PP95	Q9pp95 campylobact
27	444.5	25.5	333	16 Q9ZD93	Q9zd93 rickettsia
28	441.5	25.4	568	16 Q9RWU8	Q9rwu8 deinococcus
29	440.5	25.3	352	16 Q987A3	Q987a3 rhizobium l
30	439	25.2	400	16 Q9A2D0	Q9a2d0 caulobacter
31	433	24.9	327	16 Q98DQ4	Q98dq4 rhizobium l
32	426.5	24.5	336	16 Q8U6P8	Q8u6p8 agrobacteri
33	425.5	24.5	448	5 Q95XY8	Q95xy8 caenorhabdi
34	423	24.3	512	5 Q21080	Q21080 caenorhabdi
35	419	24.1	334	16 Q985M4	Q985m4 rhizobium l
36	414	23.8	334	16 Q92WC4	Q92wc4 rhizobium m
37	413	23.7	600	3 Q94634	Q94634 schizosacch
38	409.5	23.5	508	16 P73375	P73375 synecocyst
39	406.5	23.4	515	16 Q91418	Q91418 pseudomonas
40	399	22.9	349	17 Q9T8V1	Q9t8v1 aeropyrum p
41	398	22.9	514	16 Q8ZAB4	Q8zab4 yersinia pe
42	391	22.5	510	16 Q9KWI1	Q9kw1 vibrio chol
43	388.5	22.3	602	10 Q9AUQ1	Q9auq1 oryza sativ
44	388.5	22.3	602	10 Q8W314	Q8w314 oryza sativ
45	384	22.1	514	16 Q8Z378	Q8z378 salmonella

## ALIGNMENTS

RESULT 1:  
Q9T0D1 PRELIMINARY; PRT; 346 AA.  
AC Q9T0D1;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE T5C23.70 protein.  
GN T5C23.70 OR AT4G11640.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,  
RA Bancroft I., Mewes H.W., Mayer K.F.X., Schueller C.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,  
RA Mewes H.W., Lencic K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL049500; CAB39935.1; -  
DR EMBL; AL161532; CAB78207.1; -  
DR HSSP; P04968; 1TDJ  
DR InterPro; IPR001926; B6\_enzyme\_beta.  
DR InterPro; IPR000634; S/T\_dehydrtase.  
DR Pfam; PF00291; PALP; 1.  
DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.

```
KW Hypothetical protein.
SQ SEQUENCE 346 AA; 36839 MW; 47C37E1CA52B91F0 CRC64;

Query Match 45.5%; Score 791; DB 10; Length 346;
Best Local Similarity 46.9%; Pred. No. 7.9e-55;
Matches 159; Conservative 68; Mismatches 88; Indels 24; Gaps 6;

QY 4 QYCISPADVEKAHINQDSIHLTPVLTSSILNOIAGRNLFKCELFQ----- 50
Db 7 KYAADILSKEAHDRIKPIHRTPVLTSSLSISGRSUFFKCECIQKGECSFYFSY 66
QY 51 --KTGSKIRGALNAIRGLIPDPEEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVV 108
Db 67 CGKSGAFKRGACNAVLSL---DAEQAAKGVVTHSSGNHAAALSLAAKIQGIPAYIV 123
QY 109 TAPNCKKLAIOAYGASIVYCDPDESREKVTQIMQETEGILVHPNQEPAVTAGQTIAL 168
Db 124 GAPKCVQNDVIRYGGKVIWSEATPMSREETASKVLQETGSLVHPYNDGRITISG 183
QY 169 EVLNQVPLVDALVVPVGGGVMAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKG 228
Db 184 ELLEQIODEDAIVVPSIGGLISCVALAASKIPSIIRIATAEPKGAADAAQSKVAG 242
QY 229 NLHPPETIADGVKSSIGLNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLI 288
Db 243 TLPVTNTIADGLRASGLDTPVVRDLVDDVFTVTEDEICEIEAMKCYEILKVSPE 302
QY 289 ALAAVLSQHFQTVSP---EVKNVCIVLGGNVDLTSLNW 324
Db 303 GLAAVLSNFRN-NPSCRCKNIGIVLGGNVDLGLS-W 339

RESULT 2
Q8XPX0 PRELIMINARY; PRT; 323 AA.
AC Q8XPX0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Probable amino-acid dehydratase protein (EC 4.2.-.-).
OS RSP1516 OR RS04793.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646085; CAD18667.1; -
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF00291; PALP; 1.
KW Lyase; Plasmid; Complete proteome.
SQ SEQUENCE 323 AA; 34515 MW; 2175CBF693E04A2D CRC64;

Query Match 35.6%; Score 620; DB 16; Length 323;
Best Local Similarity 41.0%; Pred. No. 2.9e-41;
Matches 133; Conservative 65; Mismatches 118; Indels 8; Gaps 4;

QY 7 ISFADVEKAHINQDSIHLTPVLTSSILNOIAGRNLFKCELFQKTSFKIRGALNAIR 66
Db 6 IYDDVVAHARLQGVHRTPLVLTSTANALGAELFFKCECFQNGAFKFRGAYNALSQ 65
QY 67 LIPDTPPEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVQTPAPNCKKLAIOAY 126
```

```
Db 66 F---TPEQRKAGVITPSSGNHQAIALSARLLGMRAVIMPKDAPTIKVEATRGYGEV 122
QY 127 YCDPSDESREKVTQIMQETEGILVHPNQEPAVIAGQGTIALEVLNOVPLVDALV 186
Db 123 FYDRYTEDREAIGRRILAEQHGLTLIPPDYHPHYMAGQTAAKELIETGPDLLLV 182
QY 187 GNMVAGIATIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPETIADGVKSS-IG 245
Db 183 GLLSCSCATAARALSAPACRIIGVEPAGNGQSLRKEIV-HIDTPATIADGAQTHLG 241
QY 246 LNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLIETAGVALAAVLSQHFQTV 305
Db 242 HYTEFAVIRALVDDIATVSDADLVDTMRFFAGRMKIVVPTGCLAAAAAARQR---VEVRG 298
QY 306 KNCVCIVLGGNVDLTSLNWVQNAE 329.
Db 299 KRVGVIISGNVDLQHFARLVQAD 322

RESULT 3
Q8X0J0 PRELIMINARY; PRT; 388 AA.
AC Q8X0J0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Related to threonine dehydratase.
OS B11H24.140.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL670005; CAD21283.1; -
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF00291; PALP; 1.
SQ SEQUENCE 388 AA; 41505 MW; 02BFB864B95716A6 CRC64;

Query Match 31.4%; Score 546; DB 3; Length 388;
Best Local Similarity 37.5%; Pred. No. 2.9e-35;
Matches 140; Conservative 63; Mismatches 106; Indels 64; Gaps 10;

QY 10 ADVEKAHINQDSIHLTPVLTSSILNOIAGR-----NLFKFC 46
Db 13 ASVLAHSLIKPIYIHPVLTNTFDQLASTPRTPELQCTEWEATERPANPKIREWFKC 72
QY 47 ELFOKTSFKIRGALNAIRGLIPDPEE--KPKAVVTHSSGNHGQALTYAAKLEGIPAYI 104
Db 73 ENFORIGAFKARGAFHAFERLQTEGLGKKGVVVTHSSGNHQAALSAARENGIPAH 132
QY 105 VVPOTAPNCKKLAIOAYGASIVYCDPDESREKVTQIMQETEGILVHPNQEPAVIAG 164
Db 133 VMPTISPPPKIAATKYGANITFSGSTSTEREAVTREVIEKTCARLVPPYDHPDIL 192
QY 165 TALEVLNQVPL-----VDALVVPVGGGVMAGIAITIKALKPS----- 203
Db 193 TAALELQROVAASLSSTGTTTTNRRRLNAITPCGGGLLSCTALACSDLSPSDPTG 252
QY 204 -VKVYAAEP--SNADD---CYOSKLGELTPNLHPETIADGVKSSIGLNTWPII--RDL 255
Db 253 PILVGAEPFSGADGRGRGYSGTRIESVKSL---TIADGLRPLGAYPWSIYERKL 308
QY 256 VDDVFTVTEDEIKYATQLVWGRMKLLIETAGVALA-AVLSSHQFQTV-----SPEK 309
```

```

Db 309 VAGWYSGEIBEIKALRLVYRMKVVVPEPSAVGLAVLAFNEFRSMYVERGGEGWDLG 368
Qy 310 IVLSGGNVDLTSL 322
   :| |||||:| :|
Db 369 VVFSGGNVLAAL 381

RESULT 4
Q910F5 PRELIMINARY; PRT; 320 AA.
AC Q910F5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Probable serine/threonine dehydratase, degradative.
GN PA2683.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.H., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004696; AAC06071.1; -.
DR HSSP; P04968; 1TDJ.
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF00291; PALP; 1.
KW Complete proteome.
SQ SEQUENCE 320 AA; 33957 MW; C6F97C6259FE8584 CRC64;

Query Match 31.3%; Score 544; DB 16; Length 320;
Best Local Similarity 39.08; Pred. No. 3.2e-35;
Matches 122; Conservative 62; Mismatches 121; Indels 8; Gaps 4;

Qy 8 SFADVEKAHINIQDSIHLPVLTSSILNQIAGRNLFKCELFQKTSFKIRGALNAIRGL 67
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 6 TYDDVIAAARIAIAGNRPVMSRSLDEELGAEVFEKCNLQRMGAFKFGAFLSRF 65

Qy 68 IPDTPPEKPAVYTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIQAYGASIVY 127
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 66 ---SAEQRAAGVAVFSSGNHAQALSAIRLIGIPATIVMPADAPAVKIEATRGYGQVVL 122

Qy 128 CDPSDESREKVTORIMQETEGILVHPNQEPAVIAGQGTIALEVLNQPLVDALVVPVGG 187
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 123 YDRYTEQIGRDLAQRHGLTLPYDHPDLVLAGQGTAAKELFEVGPDLDAFPAPLGGG 182

Qy 188 GWVAGIAITIKALPSPVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVKSS-IGL 246
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 183 GLLSGCALRALAPACRIKGVPEAGNDGQSLRSGAIV-HIDTPQTLADGAQTQHLGN 241

Qy 247 NTFPIIRDLDVDDFTVTDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFQTVSPVK 306
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 242 LTFPLIQNVDDIILTADAEVLDMRFLEAARMKLLVEPTGCLGLAAAAQRKDEL---RGK 298

Qy 307 NVCIVLSGGNVDL 319
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 299 RVGILLSGGNVDL 311

RESULT 5
Q99U50 PRELIMINARY; PRT; 346 AA.
AC Q99U50;

```

```

DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE SAI271 protein (Threonine deaminase IlvA homolog).
GN SAV1438 OR SAI271.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OX Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003362; BAB57600.1; -.
DR EMBL; AP003133; BAB42531.1; -.
DR HSSP; P04968; 1TDJ.
DR InterPro; IPR001926; B6_enzyme_beta.
DR InterPro; IPR000634; S/T_dehydrtse.
DR Pfam; PF00291; PALP; 1.
DR TIGRFAMS; TIGR01127; ilvA_1cterm; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW Complete proteome.
SQ SEQUENCE 346 AA; 37147 MW; 4435206B7E89CC10 CRC64;

Query Match 30.7%; Score 535; DB 16; Length 346;
Best Local Similarity 37.8%; Pred. No. 1.9e-34;
Matches 122; Conservative 64; Mismatches 131; Indels 6; Gaps 4;

Qy 5 YCISFADVEKAHINIQDSIHLPVLTSSILNQ- IAGRNLFKCELFQKTSFKIRGALNA 63
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 12 HIVSLGDIIEAKASIKPFIRTPRIKLSWYLSQNTKGNVYLUKLENNMQTSFKFRGASNK 71

Qy 64 IRGLIPDTPPEKPAVYTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIQAYGA 123
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 72 INHL---SDEQKAGIICASAGNHAQVALTAKLLGIDATIVMPETAPIAKONATKGXGA 128

Qy 124 SIVYCDPSDESREKVTORIMQETEGILVHPNQEPAVIAGQGTIALEVLNQPLVDALVVP 183
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 129 KVIILKGNFNETRLYMEELAKENGMTIVHPYDDKRFVMAGQGTIGLEILDDIWNVTIVP 188

Qy 184 VGGGMVAGIAITIKALPSPVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVKSS 243
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 189 VGGGLIAGTATALKSNFNSIHIIGVQAEVHGMAESFYKRALTEH-REDSTIADGCDVK 247

Qy 244 I-GLNTWPIIRDLDVDDFTVTDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFQTVS 302
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 248 VPGEKTEVVVKHLVDEFLVSEEEIEHAMQDLQRAKIITEGAGALPTAAILSGKIDKKW 307

Qy 303 PEVKNVCIVLSGGNVDLTSLNW 325
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 308 LEGKNVVALVSGGNVDLTRVSGV 330

RESULT 6
Q9XAA4 PRELIMINARY; PRT; 325 AA.
AC Q9XAA4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative threonine dehydratase.
GN SC00821 OR SCF43A.11C.

```

```

OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyceae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K., Harris D.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Kinschi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL096837; CAB48898.1;
DR HSSP: P04968; 1TDJ.
DR InterPro: IPR001926; B6_enzyme_beta.
DR InterPro: IPR000634; S/T_dehydratase.
DR Pfam: PF00291; PALP; 1.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
SQ SEQUENCE 325 AA; 33965 MW; 4E1DE62D340489A7 CRC64;

Query Match 30.5%; Score 530; DB 16; Length 325;
Best Local Similarity 39.3%; Pred. No. 4.2e-34;
Matches 123; Conservative 54; Mismatches 128; Indels 8; Gaps 4;

QY 7 ISFADVERKAHNIODSIHLTPVLTSIILOTAGRNLFKFCELFOKTSFKIRGALNAIRG 66
Db VLTDVRSAAARIKQVARTPVLSRTLDALVGAELHVKCNQQRVGAFKRGAYNAASR 67
QY 67 LIPDTPPEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPNCKKLAIQAYGASIV 126
Db L---TPEQLARGIAAYSSGNHAQVALAARELGITAVIVMPEDAPPSKRDATAGVAGIV 124
QY 127 YCDPSDESREKVTQIMQETEGILVHPNQEPVAVTAGOQTIALEVUNQVPLVDALVVPVGG 186
Db TYDRTYGRVAVAEALADRLGTLIPPEYHPVHVIAGOGTAALELVETGELDIALVAPVG 184
QY 187 GGMVAGIATITKALKPSVKVYAAAPSNADDCYQSKLKGELTPNLHPPTIADG-VKSSIG 245
Db LGLIAGSATYAKALHPGRVIGVPEAGDDTRRSUAGRRV-SVPVPTIADGQALPTPG 243
QY 246 LNTWPIIRDLDVDDVFTTVEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFQTVSPEV 305
Db ELTFSLNRRLLDGLVLSDDSDIARDAMFAFERLKTVPLEPSGATPLAALLNGRIDALP--- 300
QY 306 KNCVICVLSGGNVD 318
Db 301 RVGVLSGGNVD 313

```

## RESULT 7

```

Q9WYJ1
ID Q9WYJ1 PRELIMINARY; PRT; 401 AA.
AC Q9WYJ1;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Threonine dehydratase catabolic.
GN TM0356.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickley E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett K.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 393:323-329(1999).
DR EMBL: AE001716; AAD35443.1;
DR HSSP: P04968; 1TDJ.
DR TIGR: TM0356;
DR InterPro: IPR002912; ACT.
DR InterPro: IPR001926; B6_enzyme_beta.
DR InterPro: IPR000634; S/T_dehydratase.
DR Pfam: PF01842; ACT; 1.
DR Pfam: PF00291; PALP; 1.
DR TIGRFAMS: TIGR01127; ilva_1cTerm; 1.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
KW Complete proteome.
SQ SEQUENCE 401 AA; 43110 MW; 4D757A8F93786180 CRC64;

```

```

Query Match 30.0%; Score 522.5; DB 16; Length 401;
Best Local Similarity 35.8%; Pred. No. 2.3e-33;
Matches 114; Conservative 72; Mismatches 123; Indels 9; Gaps 4;

QY 7 ISFADVERKAHNIODSIHLTPVLTSIILOTAGRNLFKFCELFOKTSFKIRGALNAIRG 66
Db ITLEDIKEAQRTLNKVVHRTALAYSSVSEVTGGTGYLUKMNLOKTSFKIRGAYNKIAH 61
QY 67 LIPDTPPEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPNCKKLAIQAYGASIV 126
Db L---SEERKRGVVAASAGNHAQVALAARIGIPATIVMPRYAPLSKITKTRNLGAQVI 118
QY 127 YCDPSDESREKVTQIMQETEGILVHPNQEPVAVTAGOQTIALEVUNQVPLVDALVVPVGG 186
Db LEGNIFDEAYEAALRIQEKTKGAVFVHPFNDPHVIAGOGTIGLEIMEDLPDVEVVVPPVG 178
QY 187 GGMVAGIATITKALKPSVKVYAAAPSNADDCYQSKLKGELTPNLHPPTIADGVK-SSIG 245
Db LGLIAGSATYAKALHPGRVIGVPEAGDDTRRSUAGRRV-AERVEGKPTLADGIAVKKPG 237
QY 246 LNTWPIIRDLDVDDVFTTVEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFQTVSPEV 305
Db DTETELVKKYVDENVAVNNEEIIADAILFLLEQAKVVAAGAGVAAVLNK----LDVKG 293
QY 306 KNCVICVLSGGNVDLTSLN 323
Db 294 KKAIVISGGNIDVNMID 311

```

## RESULT 8

```

Q8UBW5
ID Q8UBW5 PRELIMINARY; PRT; 324 AA.
AC Q8UBW5;

```



```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lennard N.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z80219; CAB02298.1; -.
DR HSSP; P35520; 1JBO.
DR InterPro; IPR001926; B6_enzyme_beta.
DR InterPro; IPR000634; S/T_dehydratase.
DR Pfam; PF00291; PALP; 1.
DR TIGRFAMS; TIGR01127; livA_1Cterm; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
SQ SEQUENCE 317 AA; 34688 MW; 0F0DE5C7479389E3 CRC64;

Query Match      29.2%; Score 507.5; DB 5; Length 317;
Best Local Similarity 37.7%; Pred. No. 2.5e-32;
Matches 121; Conservative 62; Mismatches 115; Indels 23; Gaps 9;

QY 7 ISFADVEKAHINIQDSIHLPVLTSSILNQIAGR--NLFFKCELFQKTSFKIRGALNAI 64
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 5 ITMEEMAAAYQRTAPFLHKTIVITSENIDKVGNGTHVLPKCEHLQKTSFKARGALNS- 63
QY 65 RGLIPDTPPEKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPOTPANCKKLAIQAYGAS 124
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 64 -AIL--AKEKNAKGMIAHSSNGHQALAWAAQKIGLPCTIVVPPNAPISKIEGHRNAN 120
QY 125 IVYCDPSDESREKVTORIMQETEGILVHPNCEPAVIAOGGTIALEVLNOV-PLVDALVVP 183
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 121 IVFCEPTVTSRESCADLTKELEYCYEYPNCVSWINGHSSVAPEILLEQVGNEDISFLS 180
QY 184 VGGGWMVAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPP-----ETIADG 239
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 181 VGGGLASSVAFLGNLRPDIEVYLQPA-----QKELSNLENGIKCPDVTLDTIADG 234
QY 240 VK--SSIGLNTWPIIRD-LVDDVFTVTEDEIKYATQLVGRMKLLIETAGVALAAVLSOH 297
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 235 VRVARVGPLCEPILKQYCAQKVISKEDEIKGLKLIWTRMKQRIEPTAALAFAGVLYHK 294
QY 298 FQTVSPVKVNCVILSGGND 318
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 295 ----PAHKRPLVILCGGND 311

RESULT 11
Q975Y3
ID Q975Y3 PRELIMINARY; PRT; 404 AA.
AC Q975Y3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative threonine dehydratase.
GN ST0295.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic

```

```

RT Crenarchaeon, Sulfolobus tokodaii strain 7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000982; BAB65265.1; -.
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF00291; PALP; 1.
DR TIGRFAMS; TIGR01127; livA_1Cterm; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 404 AA; 44034 MW; 4C4B44D776FE26B7 CRC64;

Query Match      28.9%; Score 502.5; DB 17; Length 404;
Best Local Similarity 38.9%; Pred. No. 8.9e-32;
Matches 125; Conservative 54; Mismatches 131; Indels 11; Gaps 6;

QY 5 YCISFADVEKAHINIQDSIHLPVLTSSILNQIAGRNLFFKCELFQKTSFKIRGALNAI 64
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 3 YTKYFDEIVKIQDRIKQYIHETPIDYKTSFSDMIGAQIYKLENLQKTSFKVKGATSKL 62
QY 65 RGLIPDTPPEKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPOTPANCKKLAIQAYGAS 124
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 63 TSL---AEEERKKGVIAVSAGNHAQGVAAKVLGIRKATIVMPETAPISKYQATKSYGAQ 119
QY 125 IVYCDPSDESREKVTORIMQETEGILVHPNCEPAVIAOGGTIALEVLNOVPLVDALVVP 184
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 120 VILYKGFHESMKKAEELIREEGSILVHPYGDLDVILGGTGLGUELLPYNP--DIVVVP 177
QY 185 GGGGWMVAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPETIADG--VK 242
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 178 GGGGLSIGSIALAKAKNPKTKIGVQSSASPSLAKSKDLHLRV-EIEPFSIADGILVKS 236
QY 243 SIGLNTWPIIRDVDDVFTVTEDEIKYATQLVGRMKLLIETAGVALAAVLSOHFQTVS 302
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 237 PSEI-TFSIIDEIVLVDDEETIANAVYLLERNKTLVEGAAASLALLSGKIKVAY 295
QY 303 PEVKNVCIVLGGNVDLTSLN 323
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 296 NH--KAIPVLGGNIDLSMLS 314

RESULT 12
Q8XL77
ID Q8XL77 PRELIMINARY; PRT; 402 AA.
AC Q8XL77;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Threonine dehydratase.
GN THD OR CPE1165.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003189; BAB80871.1; -.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001926; B6_enzyme_beta.
DR InterPro; IPR000634; S/T_dehydratase.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00291; PALP; 1.
DR TIGRFAMS; TIGR01127; livA_1Cterm; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW Complete proteome.
SQ SEQUENCE 402 AA; 43421 MW; E872AC5FE71BF809 CRC64;

Query Match      28.8%; Score 501; DB 16; Length 402;
Best Local Similarity 36.8%; Pred. No. 1.2e-31;

```



```
QY 243 SI-GLNTWPIIROLVDDVFTVTDEIKYATOLVWGRMKLLIEPTAGVALAAVLSQLHFTV 301
Db 240 PIPGDLTWPINOKNLGSLVAVATDAEAMRYAFSTLKLVEPGCGVALTAALTGKVDVA 299
QY 302 SPEVKNVCIVLSGNDV 318
Db 300 G---KTVAIVLSGNDV 313

RESULT 15
Q9HNH6 PRELIMINARY; PRT; 495 AA.
AC Q9HNH6;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Threonine dehydratase.
GN ILDA OR VNG2100G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lesky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005099; AGC20244.1; -;
DR HSSP; P04968; ITDJ.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001926; B6_enzyme.beta.
DR InterPro; IPR000634; S/T_dehydrtse.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00291; PALP; 1.
DR TIGRFAMS; TIGR01127; ilva_lcterm; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 495 AA; 53095 MW; 8F203680AAF93F80 CRC64;

Query Match 28.3%; Score 492.5; DB 17; Length 495;
Best Local Similarity 36.5%; Pred. No. 7.4e-31;
Matches 119; Conservative 59; Mismatches 129; Indels 19; Gaps 7;

QY 7 ISFADVEKAHINIODSIHLTPVLTSSILNOIAGRNLFKCELFQKTSFKIRGALNAIRG 66
Db 87 LSLSDIRDAERVSETHRTPLEYSHTFSDLTGADVRLKLECFQKTSFKIRGATNRIRT 146
QY 67 LIPDTPPEKPKAVYTHSSNGHQALTYAAKLEGIPAYIVVPQTPAPNCKKLAIOAGSIV 126
Db 147 LSAD---QQDAGVVVTASAGNHAQVALAASRSQVDSKVVPESAPISKIKATKSYCAEV 203
QY 127 -----YCDPSDESREKVTQIMQETEG-ILVHPNQEPAVIAGQGTIALEVLNOVPLVDAL 180
Db 204 LHGADYDDAQAHAE-----LEAAEGRTYVHAFDDDEYIMAGQGTGLGLEIAADCPVTDV 257
QY 181 VVPVGGGVAGIATIKALKPSVKVYAAEPSNADDCYOSKLGKELTLPNHPPTIADGV 240
Db 258 VVPIGGGLISGVATALKGELDDVRVIGVQAECASTVARSLDKGQQAQVDH-VDTIADGI 316
QY 241 K-SSIGLNTWPIIROLVDDVFTVTEDEIKYATOLVWGRMKLLIEPTAGVALAAVLSQLH 299
Db 317 AVROVGAQTFPVIQEHVDEWTVVSDDEIATALLVLLERCKTLVEGAGATALAALVLEDKFE 376
QY 300 TVSPVEKNVCIVLSGNDVLSLNV 325
```

Db 377 YADGE--TIVPALCGGNIDNLTLTV 400

Search completed: June 24, 2003, 06:21:09

Job time : 45.4389 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:35:44 ; Search time 45.9323 Seconds  
(without alignments)  
983.448 Million cell updates/sec

Title: US-09-889-609B-8  
Perfect score: 1740  
Sequence: 1 MCAQYCISPADVEKAHINIQ.....TSLNWNQGAERPAPYQTVSV 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1740	100.0	339	21	AA807731 Amino acid sequenc
2	1582.5	90.9	340	21	AA807734 Amino acid sequenc
3	1582.5	90.9	340	22	AA009124 Human Serine Racem
4	1582.5	90.9	340	22	AA508342 Human pyridoxal-ph
5	1582.5	90.9	340	22	AA594477 Human protein sequ
6	1578.5	90.7	340	22	AA50262 Human serine racem
7	1574.5	90.5	340	22	AA50262 Serine/threonine d
8	988	56.8	228	23	ABB89713 Human polypeptide
9	453	26.0	469	22	ABB63708 Drosophila melanog
10	435.5	25.0	1181	22	ABG24298 Novel human diagno

11	371	21.3	502	20	AA532941 Mutant threonine d
12	371	21.3	502	20	AA505705 Feedback insensiti
13	371	21.3	532	20	AA532943 Mutant threonine d
14	371	21.3	532	20	AA505707 Feedback insensiti
15	371	21.3	539	20	AA532942 Mutant threonine d
16	371	21.3	539	20	AA505706 Feedback insensiti
17	371	21.3	545	20	AA532947 Mutant threonine d
18	371	21.3	545	20	AA505711 Feedback insensiti
19	371	21.3	592	20	AA532939 Mutant threonine d
20	371	21.3	592	20	AA532951 Wild type threonin
21	371	21.3	592	20	AA505702 Arabidopsis wild-t
22	371	21.3	592	20	AA505703 Feedback insensiti
23	371	21.3	600	20	AA532952 Mutant threonine d
24	371	21.3	609	20	AA532940 Mutant threonine d
25	371	21.3	609	20	AA505704 Feedback insensiti
26	366	21.0	590	20	AA532950 Mutant threonine d
27	366	21.0	592	20	AA532948 Mutant threonine d
28	365	21.0	424	22	AA581976 S. epidermidis ope
29	365	21.0	424	23	ABB39031 Staphylococcus epi
30	362.5	20.8	416	23	ABB54559 Lactococcus lactis
31	359.5	20.7	441	15	AA54223 L.lactis branched
32	357.5	20.5	310	22	AA50837 C glutamicum prote
33	357.5	20.5	310	22	AA579703 Corynebacterium gl
34	344.5	19.8	422	23	ABB48174 Listeria monocytog
35	325	18.7	621	22	ABG30299 Novel human diagno
36	311	17.9	316	22	ABB66014 Drosophila melanog
37	307	17.6	436	22	AA52074 C glutamicum prote
38	305	17.5	436	16	AA54590 Threonine dehydrat
39	304	17.5	423	22	AA579762 Corynebacterium gl
40	303	17.4	436	16	AA54686 Threonine dehydrat
41	302	17.4	436	16	AA54687 Threonine dehydrat
42	302	17.4	436	16	AA54688 Threonine dehydrat
43	300	17.2	436	16	AA54689 Threonine dehydrat
44	300	17.2	436	16	AA54691 Threonine dehydrat
45	282.5	16.2	340	22	AA523764 Novel human enzyme

ALIGNMENTS

RESULT 1  
AA807731  
ID AAB07731 standard; Protein; 339 AA.  
XX AC  
XX AAB07731;  
XX  
XX 07-NOV-2000 (first entry)  
XX  
XX Amino acid sequence of a murine serine racemase polypeptide.  
DE  
DE Serine racemase; N-methyl-D-aspartate receptor; neural death;  
XX  
XX neural dysfunction; NMDA receptor; Parkinson's disease;  
KW  
KW Huntington's disease; motor neurone disease; Alzheimer's disease.  
XX  
XX Mus musculus.  
OS  
XX  
XX Key Location/Qualifiers  
FH  
FH Misc-difference 337  
FT  
FT /note= "Ser encoded by T in AAA59294"  
XX  
XX WO200043526-A1.  
XX  
XX 27-JUL-2000.  
XX  
XX 18-JAN-2000; 2000WO-US00938.  
XX  
XX 19-JAN-1999; 99US-0116333.  
PR  
PR 21-JUL-1999; 99US-0144839.  
PR  
PR 28-JUL-1999; 99US-0145953.  
XX  
XX (UOJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
PA  
PA Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;  
XX  
PI

```

PI Ferris CD;
XX
DR WPI; 2000-482915/42.
DR N-PSDB; AAA59294, AAA59300.
XX
XX Mammalian serine racemase preparations, used to identify modulators
PT which can be used to treat diseases associated with
PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease
PT
XX
XX Claim 7; Page 48-49; 54pp; English..
PS
XX
XX The present sequence represents a mammalian serine racemase, which has
CC a specific activity of at least 0.003 micromole L-serine/mg/hour.
CC The enzyme catalyses the direct racemisation of L-serine to D-serine.
CC D-serine appears to be an endogenous ligand of N-methyl-D-aspartate
CC (NMDA) receptors. The mammalian serine racemases can be used to identify
CC modulators, which can be used in the treatment of acute or chronic
CC neural death or dysfunction mediated by overactivation of N-methyl-D-
CC aspartate (NMDA) receptors. Overactivation of the receptors is
CC associated with Parkinson's disease, Huntington's disease, motor neurone
CC disease and Alzheimer's disease.
XX
XX Sequence 339 AA;
SQ
Query Match 100.0%; Score 1740; DB 21; Length 339;
Best Local Similarity 100.0%; Pred. No. 2.7e-169;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MCAQYICISFADVEKAHINIQDSIHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGA 60
DB 1 MCAQYICISFADVEKAHINIQDSIHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGA 60
QY 61 LNAIRGLIPTPEKPKAVVTHSSGNHGQALTYAAKLEGIPAVIVVQTPAPNCKKLAIOA 120
DB 61 LNAIRGLIPTPEKPKAVVTHSSGNHGQALTYAAKLEGIPAVIVVQTPAPNCKKLAIOA 120
QY 121 YGASIVVCDPSDESREKVTQRIHQETEGILVHPNQEPVAVIAGOGTTIALEVLNQVPLVDAL 180
DB 121 YGASIVVCDPSDESREKVTQRIHQETEGILVHPNQEPVAVIAGOGTTIALEVLNQVPLVDAL 180
QY 181 VVPVGGGWVAGIATITKALKPSVKVYAAEPSNADDCYOSKLGELTPNLHPPTIADGV 240
DB 181 VVPVGGGWVAGIATITKALKPSVKVYAAEPSNADDCYOSKLGELTPNLHPPTIADGV 240
QY 241 KSSIGLNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFTQ 300
DB 241 KSSIGLNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFTQ 300
QY 301 VSPVKNVCIVLSSGGNVDLTSLNWVGQAERPAPYQTVSV 339
DB 301 VSPVKNVCIVLSSGGNVDLTSLNWVGQAERPAPYQTVSV 339
RESULT 2
AAB07734
ID AAB07734 standard; Protein; 340 AA.
XX
XX AAB07734;
XX
XX 07-NOV-2000 (first entry)
XX
XX Amino acid sequence of a human serine racemase polypeptide.
XX
XX Serine racemase; N-methyl-D-aspartate receptor; neural death;
XX neural dysfunction; NMDA receptor; Parkinson's disease;
XX Huntington's disease; motor neurone disease; Alzheimer's disease.
XX
XX Homo sapiens.
XX
XX W0200043526-A1.
XX
XX 27-JUL-2000.
XX

```

---

```

XX
PF 18-JAN-2000; 2000WO-US00938.
XX
PR 19-JAN-1999; 99US-0116333.
PR 21-JUL-1999; 99US-0144839.
PR 28-JUL-1999; 99US-0145953.
XX
XX (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
XX Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;
PI Ferris CD;
XX
XX WPI; 2000-482915/42.
DR N-PSDB; AAA59299.
XX
XX Mammalian serine racemase preparations, used to identify modulators
PT which can be used to treat diseases associated with
PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease
PT
XX
XX Claim 8; Page 50-51; 54pp; English.
PS
XX
XX The present sequence represents a mammalian serine racemase, which has
CC a specific activity of at least 0.003 micromole L-serine/mg/hour.
CC The enzyme catalyses the direct racemisation of L-serine to D-serine.
CC D-serine appears to be an endogenous ligand of N-methyl-D-aspartate
CC (NMDA) receptors. The mammalian serine racemases can be used to identify
CC modulators, which can be used in the treatment of acute or chronic
CC neural death or dysfunction mediated by overactivation of N-methyl-D-
CC aspartate (NMDA) receptors. Overactivation of the receptors is
CC associated with Parkinson's disease, Huntington's disease, motor neurone
CC disease and Alzheimer's disease.
XX
XX Sequence 340 AA;
SQ
Query Match 90.9%; Score 1582.5; DB 21; Length 340;
Best Local Similarity 89.7%; Pred. No. 3.7e-153;
Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;
QY 1 MCAQYICISFADVEKAHINIQDSIHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGA 60
DB 1 MCAQYICISFADVEKAHINIQDSIHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGA 60
QY 61 LNAIRGLIPTPEKPKAVVTHSSGNHGQALTYAAKLEGIPAVIVVQTPAPNCKKLAIOA 120
DB 61 LNAIRGLIPTPEKPKAVVTHSSGNHGQALTYAAKLEGIPAVIVVQTPAPNCKKLAIOA 120
QY 121 YGASIVVCDPSDESREKVTQRIHQETEGILVHPNQEPVAVIAGOGTTIALEVLNQVPLVDAL 180
DB 121 YGASIVVCDPSDESREKVTQRIHQETEGILVHPNQEPVAVIAGOGTTIALEVLNQVPLVDAL 180
QY 181 VVPVGGGWVAGIATITKALKPSVKVYAAEPSNADDCYOSKLGELTPNLHPPTIADGV 240
DB 181 VVPVGGGWVAGIATITKALKPSVKVYAAEPSNADDCYOSKLGELTPNLHPPTIADGV 240
QY 241 KSSIGLNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFTQ 300
DB 241 KSSIGLNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFTQ 300
QY 301 VSPVKNVCIVLSSGGNVDLTSLNWVGQAERPAPYQTVSV 339
DB 301 VSPVKNVCIVLSSGGNVDLTSLNWVGQAERPAPYQTVSV 339
RESULT 3
AAB09124
ID AAB09124 standard; Protein; 340 AA.
XX
XX AAB09124;
XX
XX 16-JAN-2002 (first entry)
XX
XX Human Serine Racemase.
XX

```

XX Human; serine racemase; Parkinson's disease; Huntington's disease;  
KW anxiety; glaucoma; stroke; hyperalgesia; pain; spinocerebellar ataxia;  
KW schizophrenia; transgenic animal; chromosome 17p13; immunogen.  
XX Homo sapiens.  
PN WO200175144-A1.  
XX 11-OCT-2001.  
XX 02-APR-2001; 2001WO-US10662.  
PF 04-APR-2000; 2000US-194451P.  
XX (MERI ) MERCK & CO INC.  
XX Connolly T, Liu Y, Xia M;  
PI WPI; 2001-656991/75.  
XX N-PSDB; AAS15217.  
XX New recombinant serine racemase polypeptide, useful in assays for  
PT identifying compounds that alter enzyme activity (e.g. including  
PT compounds that inhibit or stimulate enzyme activity) or in generating  
PT antibodies against the protein  
XX Claim 7; Page 14; 43pp; English.  
XX The invention relates to a recombinant polynucleotide encoding a  
CC human serine racemase, vectors containing it, host cells expressing the  
CC racemase, methods of identifying inhibitors of serine racemase and a  
CC transgenic animal lacking a functional endogenous serine racemase  
CC comprising the human serine racemase of the invention. The racemase  
CC protein is useful in assays to identify compounds that inhibit or  
CC stimulate enzyme activity, in the generation of antibodies against the  
CC protein, and in structural studies of the protein and  
CC structure/function relationships of the protein. Biologically active  
CC polypeptide have diagnostic, therapeutic or prophylactic uses (e.g.  
CC for neurological diseases such as Parkinson's and Huntington's disease,  
CC anxiety, glaucoma, stroke, hyperalgesia, pain, spinocerebellar ataxia and  
CC schizophrenia), and would be useful for screening for modulators and/or  
CC inhibitors of serine racemase function. The polynucleotides are useful  
CC as probes for the specific detection of the presence of a polynucleotide  
CC encoding a serine racemase protein, and as primers for nucleic acid  
CC amplification based assays for the detection of polynucleotides encoding  
CC serine racemase protein. The transgenic animal is useful for the study of  
CC the tissue and temporal specific expression or activity of the serine  
CC racemase gene in an animal. The gene for human serine racemase is  
CC located on chromosome 17p13. The present sequence represents human serine  
CC racemase.  
XX Sequence 340 AA;  
SQ

Query Match 90.9%; Score 1582.5; DB 22; Length 340;  
Best Local Similarity 89.7%; Pred. No. 3.7e-153;  
Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;  
QY 1 MCAQYICISFADVEKAHINTQDSIHLPPVLTSSILNQIAGRNLFKCELFQKTSFKIRGA 60  
DB 1 MCAQYICISFADVEKAHINTQDSIHLPPVLTSSILNQIAGRNLFKCELFQKTSFKIRGA 60  
QY 61 LNAIRGLIPDTPPEKPAVVTHSSGNHGOALTYAAKLEGIPAYIVVPOTPAPNCKKLAIOA 120  
DB 61 LNAIRSLVPDALEKPKAVVTHSSGNHGOALTYAAKLEGIPAYIVVPOTPAPNCKKLAIOA 120  
QY 121 YGASIVYCDPSDESREKVTQIRMQETEGILVHPNQEPVIAIGQTTALEVLNOVPLVDAL 180  
DB 121 YGASIVYCDPSDESRENKRVTEETEGILVHPNQEPVIAIGQTTALEVLNOVPLVDAL 180  
QY 181 VVPVGGGGVAGIATITKALKPSVKVYAAEPSNADDCYOSKLGKLPNLYPPETIADGV 240  
DB 181 VVPVGGGGVAGIATITKALKPSVKVYAAEPSNADDCYOSKLGKLPNLYPPETIADGV 240

DB 181 VVPVGGGGVAGIATITKALKPSVKVYAAEPSNADDCYOSKLGKLPNLYPPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAVALSOHQFOT 300  
DB 241 KSSIGLNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAVALSOHQFOT 300  
QY 301 VSPEVKNVCIVLSSGGNVDLT-SLNWVGQAERPAPYQTVSV 339  
DB 301 VSPEVKNVCIVLSSGGNVDLTSSITWVKQERPASYQSVSV 340  
RESULT 4  
AAE08342  
ID AAE08342 standard; Protein; 340 AA.  
XX AC AAE08342;  
XX 15-NOV-2001 (first entry)  
XX Human pyridoxal-phosphate dependent enzyme 22406 protein.  
XX Human; pyridoxal phosphate dependent enzyme; nootropic; neuroprotective;  
KW anticonvulsant; cerebroprotective; cardiant; vasotropic; gene therapy;  
KW epilepsy; Alzheimer's disease; chromosome 17; serine racemase; stroke;  
KW behavioural change; neurodegenerative disorder; schizophrenia; atresia;  
KW rheumatic heart failure; circulatory disorder; hepatic injury; jaundice;  
KW lung disorder; nodular hyperplasia; stenosis; skeletal muscle disorder;  
KW tumour; rhabdomyosarcoma; dermal fibroblast disorder.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FT Modified-site 8..11  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Domain 19..315  
FT /note= "Pyridoxal-phosphate dependent enzyme family  
FT domain"  
FT Modified-site 38..40  
FT /label= Protein\_kinase\_C\_phosphorylation\_site  
FT Modified-site 47..60  
FT /note= "Serine/threonine dehydratases pyridoxal-phosphate  
FT attachment site"  
FT Modified-site 54..56  
FT /label= Protein\_kinase\_C\_phosphorylation\_site  
FT Modified-site 59..64  
FT /label= N\_myristoylation\_site  
FT Modified-site 88..93  
FT /label= N\_myristoylation\_site  
FT Modified-site 109..112  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Modified-site 140..143  
FT /note= "cAMP and cGMP-dependent protein kinase  
FT phosphorylation site"  
FT Domain 176..197  
FT /label= Transmembrane\_domain  
FT Modified-site 187..192  
FT /label= N\_myristoylation\_site  
FT Modified-site 196..198  
FT /label= Protein\_kinase\_C\_phosphorylation\_site  
FT Modified-site 203..205  
FT /label= Protein\_kinase\_C\_phosphorylation\_site  
FT Modified-site 212..215  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Modified-site 235..238  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Modified-site 239..244  
FT /label= N\_myristoylation\_site  
FT Modified-site 261..264  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Modified-site 287..292  
FT /label= N\_myristoylation\_site  
FT Domain 308..326  
FT /label= Transmembrane\_domain

XX WO200160987-A1.  
PN 23-AUG-2001.  
XX 20-FEB-2001; 2001WO-US05365.  
XX 17-FEB-2000; 2000US-0183208.  
XX (MILL-) MILLENNIUM PHARM INC.  
PA Meyers RA, Rudolph-Owen LA;  
XX WPI; 2001-529909/58.  
XX N-PSDB; AAD14461.  
XX Novel polypeptide of the human pyridoxal phosphate dependent family  
PT useful in screening and detection assays and for treatment, e.g. of  
PT epilepsy and Alzheimer's  
XX Claim 8; Fig 1; 121pp; English.  
XX The present sequence is human pyridoxal phosphate dependent enzyme  
CC 22406 which is a serine racemase. Human 22406 gene is located on  
CC chromosome 17 between D17S849 and D17S796. The protein 22406  
CC is a modulator of D-serine. D serine has been shown to modify  
CC behavioural changes associated with learning, memory and convulsions.  
CC Human 22046 and compounds that modulate the expression or activity are  
CC used to treat or diagnose neurodegenerative disorders including  
CC Alzheimer's disease, schizophrenia as well as quell anxiety and  
CC epilepsy and prevent damage from stroke as well as cardiac (heart  
CC failure, rheumatic heart failure) and circulatory disorders, liver  
CC disorders (hepatic injury, jaundice), lung disorders, prostate  
CC disorders (benign enlargement, nodular hyperplasia), colon disorders  
CC (atresia, stenosis), skeletal muscle disorders (tumours-rhabdomyosarcoma)  
CC and dermal fibroblast disorders. Human 22406 cDNA is also useful in gene  
CC therapy.  
XX  
SQ Sequence 340 AA;  
Query Match 90.9%; Score 1582.5; DB 22; Length 340;  
Best Local Similarity 89.7%; Pred. No. 3.7e-153;  
Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;  
QY 1 MCAQYICISFADVEKAHINIQDSIHLTPVLTSSILNOIAGRNLFKCELFQKTSFKIRGA 60  
Db 1 MCAQYICISFADVEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGA 60  
QY 61 LNAIRGLIPDTPPEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPNCKKLAIOA 120  
Db 61 LNAVSLVPDALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPNCKKLAIOA 120  
QY 121 YGASIVYCDPSDESREKVTORIMQETGILVHPNQEPVAVTAGOGTIALEVLNQVPLVDAL 180  
Db 121 YGASIVYCPDESRENKRVTEETEGIVHPNQEPVAVTAGOGTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGVAGTAITIKALKPSKVYAAEPSNADDCYOSKLGELTPNLHPPETIADGV 240  
Db 181 VVPVGGGGLAGIAITVFKALKPSKVYAAEPSNADDCYOSKLGKLPNLYPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLVDVFTVTEDEIKYATQLVWGRMKLLIETPTAGVALAVALSHQFOT 300  
Db 241 KSSIGLNTWPIIRDLVDDIFTVTEDEIKATQLVWERMKLLIETPTAGVGAVALSHQFOT 300  
QY 301 VSPVKNVICVLSSGGNVDLT-SLNWVGQARPAVYQTVSV 339  
Db 301 VSPVKNVICVLSSGGNVDLTSSITWVQARERASVQSVSV 340  
RESULT 5  
AAB94477  
ID AAB94477 standard; Protein; 340 AA.  
XX

AC AAB94477;  
XX 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:15149.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
XX full-length cDNAs defined in the specification, and for the detection  
XX and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX Claim 8; SEQ ID 15149; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesising 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
XX to the complementary strand of a polynucleotide which comprises one of  
XX the 5602 nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence, where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in  
XX the specification. The primer sets can be used in antisense therapy and  
XX in gene therapy. The primers are useful for synthesising polynucleotides,  
XX particularly full-length cDNAs. The primers are also useful for the  
XX detection and/or diagnosis of the abnormality of the proteins encoded by  
XX the full-length cDNAs. The primers allow obtaining of the full-length  
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
XX AAH95893 represent human amino acid sequences; and AAH13629 to  
XX AAH13632 represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.  
XX Sequence 340 AA;  
SQ  
Query Match 90.9%; Score 1582.5; DB 22; Length 340;  
Best Local Similarity 89.7%; Pred. No. 3.7e-153;  
Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;  
QY 1 MCAQYICISFADVEKAHINIQDSIHLTPVLTSSILNOIAGRNLFKCELFQKTSFKIRGA 60  
Db 1 MCAQYICISFADVEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGA 60  
QY 61 LNAIRGLIPDTPPEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPNCKKLAIOA 120  
Db 61 LNAVSLVPDALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPNCKKLAIOA 120  
QY 121 YGASIVYCDPSDESREKVTORIMQETGILVHPNQEPVAVTAGOGTIALEVLNQVPLVDAL 180  
Db 121 YGASIVYCPDESRENKRVTEETEGIVHPNQEPVAVTAGOGTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGVAGTAITIKALKPSKVYAAEPSNADDCYOSKLGELTPNLHPPETIADGV 240  
Db 181 VVPVGGGGLAGIAITVFKALKPSKVYAAEPSNADDCYOSKLGKLPNLYPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLVDVFTVTEDEIKYATQLVWGRMKLLIETPTAGVALAVALSHQFOT 300  
Db 241 KSSIGLNTWPIIRDLVDDIFTVTEDEIKATQLVWERMKLLIETPTAGVGAVALSHQFOT 300  
QY 301 VSPVKNVICVLSSGGNVDLT-SLNWVGQARPAVYQTVSV 339  
Db 301 VSPVKNVICVLSSGGNVDLTSSITWVQARERASVQSVSV 340  
RESULT 5  
AAB94477  
ID AAB94477 standard; Protein; 340 AA.  
XX

QY 121 YGASIVYCDPSDESREKVTORIMQETEGILVHPNQEPVIAAGQTTIALEVLNQVPLVDAL 180  
DB 121 YGASIVYCDPSDESREKVTORIMQETEGILVHPNQEPVIAAGQTTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGVMAGTAITIKALKPSVKVYAAEPSNADDCYQSKLGELTPNLPETIADGV 240  
DB 181 VVPVGGGVMAGTAITIKALKPSVKVYAAEPSNADDCYQSKLGELTPNLPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHOFQ 300  
DB 241 KSSIGLNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHOFQ 300  
QY 301 VSPEVKNVICVLSGGNVLDLT-SLNWVQQAERPAPYQTVSV 339  
DB 301 VSPEVKNVICVLSGGNVLDLTSSITWVKQAERPASYQSVSV 340

RESULT 6  
AAM50262  
ID AAM50262 standard; Protein: 340 AA.  
XX AC AAM50262;  
XX DT 21-JAN-2002 (first entry)  
XX DE Human serine racemase.  
XX KW Serine racemase; human; D-serine; regulation;  
KW glutamate N-methyl-D-aspartate receptor; neurodegenerative disease;  
KW stroke; Alzheimer's disease; Parkinson's disease; pain; myopathy;  
KW neurotrophic; neuroprotective; cerebroprotective; antiparkinsonian;  
KW analgesic; diagnosis; therapy; screening.  
XX OS Homo sapiens.  
XX FT Key Location/Qualifiers  
FT Modified-site 41..59  
FT /note= "prosite serine/threonine dehydratase  
pyridoxal-phosphate attachment site"  
FT Modified-site 71..128  
FT /note= "serine/threonine dehydratase  
pyridoxal-phosphate attachment site"  
FT Modified-site 154..169  
FT /note= "cysteine synthase/cystathione beta-synthase  
attachment site"  
XX PN WO200173077-A2.  
XX PD 04-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-EP03668.  
XX PR 31-MAR-2000; 2000US-193748P.  
XX PR 03-APR-2000; 2000US-194249P.  
XX PA (FARB ) BAYER AG.  
XX PI Ramakrishnan; S;  
XX DR WPI; 2001-648444/74.  
XX DR N-PSDB; AAI70575.

XX Polynucleotide encoding serine racemase enzyme and the enzyme useful  
PT for screening reagents regulating the activity of the enzyme in a  
PT neuron disease caused by over- or under-activation of glutamate  
PT N-methyl-D-aspartate  
XX Claim 1; Fig 2; 66pp; English.  
XX PS  
XX CC The present sequence is that of human serine racemase, an enzyme  
CC that catalyses the conversion of L-serine to D-serine. Neuron  
CC damage following various nervous system diseases is often caused by  
CC activation of glutamate N-methyl-D-aspartate (NMDA) receptors in

CC the brain. This receptor is activated by the binding of D-serine.  
CC Regulation of D-serine levels through regulation of serine racemase  
CC may therefore prevent or minimise neuron damage in neurogenic and  
CC myopathic disorders, neurodegenerative disorders such as Alzheimer's  
CC disease and Parkinson's disease, and disorders leading to peripheral  
CC and chronic pain. The invention provides serine racemase  
CC polypeptides and methods for their recombinant production. These  
CC polypeptides are used in claimed methods of screening for agents  
CC that modulate or decrease serine racemase activity. A claimed  
CC pharmaceutical composition comprising either an expression vector  
CC that contains a serine racemase polynucleotide, or a reagent that  
CC modulates serine racemase enzyme activity, is used to modulate  
CC serine racemase activity in a disease, particularly neuron damage or  
CC a neurodegenerative disease caused by the over- or under-activation  
CC of the glutamate NMDA receptor.  
XX  
SQ Sequence 340 AA;

Query Match 90.7%; Score 1578.5; DB 22; Length 340;  
Best Local Similarity 89.4%; Pred. No. 9.4e-153;  
Matches 304; Conservative 20; Mismatches 15; Indels 1; Gaps 1;  
QY 1 MCAQYCIISFADVEKAKHINIRDSIHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGA 60  
DB 1 MCAQYCIISFADVEKAKHINIRDSIHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGA 60  
QY 61 LNAIRGLIPDTPEEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPNCKKLAIOA 120  
DB 61 LNAVRSLLPDALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPNCKKLAIOA 120  
QY 121 YGASIVYCDPSDESREKVTORIMQETEGILVHPNQEPVIAAGQTTIALEVLNQVPLVDAL 180  
DB 121 YGASIVYCDPSDESREKVTORIMQETEGILVHPNQEPVIAAGQTTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGVMAGTAITIKALKPSVKVYAAEPSNADDCYQSKLGELTPNLPETIADGV 240  
DB 181 VVPVGGGVMAGTAITIKALKPSVKVYAAEPSNADDCYQSKLGELTPNLPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHOFQ 300  
DB 241 KSSIGLNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHOFQ 300  
QY 301 VSPEVKNVICVLSGGNVLDLT-SLNWVQQAERPAPYQTVSV 339  
DB 301 VSPEVKNVICVLSGGNVLDLTSSITWVKQAERPASYQSVSV 340

RESULT 7  
AAG78808  
ID AAG78808 standard; Protein: 340 AA.  
XX AC AAG78808;  
XX DT 27-NOV-2001 (first entry)  
XX DE Serine/threonine dehydratase 37.  
XX KW Serine/threonine dehydratase 37; cancer; HIV infection; anti-HIV;  
XX KW Cytostatic.  
XX OS Unidentified.  
XX PN CN1300824-A.  
XX PD 27-JUN-2001.  
XX PF 21-DEC-1999; 99CN-0125662.  
XX PR 21-DEC-1999; 99CN-0125662.  
XX PA (UYFU-) UNIV FUDAN.  
XX PI Mao Y, Xie Y;

```

XX WPI; 2001-530468/59.
DR N-PSDB; AAI65020.
XX
PT Polypeptide-serine/threonine dehydrase 37 and polynucleotide for coding
PT this polypeptide -
XX
XX Claim 1; Page 27 (Disclosure); 34pp; Chinese.
XX
CC The present sequence is the protein sequence for serine/threonine
CC dehydrase 37. The dehydrase and its coding sequence are useful for
CC treating diseases e.g. cancer and HIV infection.
XX
SQ Sequence 340 AA;
Query Match 90.5%; Score 1574.5; DB 22; Length 340;
Best Local Similarity 89.4%; Pred. No. 2.4e-152;
Matches 304; Conservative 19; Mismatches 16; Indels 1; Gaps 1;
QY 1 MCAQYICISFADVEKAHINIQDSIHLPVLTSSILNQIAGRNLFKCFELFKTGSFKIRGA 60
DB 1 MCAQYICISFADVEKAHINIRDSIHLPVLTSSILNQIAGRNLFKCFELFKTGSFKIRGA 60
QY 61 LNAIRGLIPTPEEKPKAVVTHSSGNHGQALTYAAKLEGIPAVIVVPTAPNCKKLAIOA 120
DB 61 LNAIRGLIPTPEEKPKAVVTHSSGNHGQALTYAAKLEGIPAVIVVPTAPNCKKLAIOA 120
QY 121 YGASIVYCDPSDESREKVTQRIHQETEGILVHPNOEPAVITAGOTTALEVLNOVPLVDAL 180
DB 121 YGASIVYCDPSDESREKVTQRIHQETEGILVHPNOEPAVITAGOTTALEVLNOVPLVDAL 180
QY 181 VVPVGGGMVAGIATITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTADGV 240
DB 181 VVPVGGGMVAGIATITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTADGV 240
QY 241 KSSIGLNTWPIIRDLDVDDVTTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFT 300
DB 241 KSSIGLNTWPIIRDLDVDDVTTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFT 300
QY 301 VSPEVKNVICVLSSGNNVLDLT-SLNWVGOAERPAPYQTVSV 339
DB 301 VSPEVKNVICVLSSGNNVLDLTSSITWVKQERPAVSQSVSV 340
XX
RESULT 8
ABB89713
ID ABB89713 standard; Protein; 228 AA.
XX
XX ABB89713;
XX
XX 24-MAY-2002 (first entry)
XX
XX Human polypeptide SEQ ID NO 2089.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.
XX
XX Homo sapiens.
XX
XX WO200190304-A2.
XX
XX 29-NOV-2001.
XX
XX 18-MAY-2001; 2001WO-US16450.
XX
XX 19-MAY-2000; 2000US-205515P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;

```

```

XX WPI; 2002-122018/16.
DR N-PSDB; ABL90122.
XX
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
XX prevention of neural, immune system, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative
XX disorders -
XX
XX Claim 11; SEQ ID NO 2089; 2081pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
XX (ABB9040-ABB90444) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 228 AA;
Query Match 56.8%; Score 988; DB 23; Length 228;
Best Local Similarity 89.8%; Pred. No. 1.5e-92;
Matches 189; Conservative 12; Mismatches 10; Indels 0; Gaps 0;
QY 1 MCAQYICISFADVEKAHINIQDSIHLPVLTSSILNQIAGRNLFKCFELFKTGSFKIRGA 60
DB 1 MCAQYICISFADVEKAHINIRDSIHLPVLTSSILNQIAGRNLFKCFELFKTGSFKIRGA 60
QY 61 LNAIRGLIPTPEEKPKAVVTHSSGNHGQALTYAAKLEGIPAVIVVPTAPNCKKLAIOA 120
DB 61 LNAIRGLIPTPEEKPKAVVTHSSGNHGQALTYAAKLEGIPAVIVVPTAPNCKKLAIOA 120
QY 121 YGASIVYCDPSDESREKVTQRIHQETEGILVHPNOEPAVITAGOTTALEVLNOVPLVDAL 180
DB 121 YGASIVYCDPSDESREKVTQRIHQETEGILVHPNOEPAVITAGOTTALEVLNOVPLVDAL 180
QY 181 VVPVGGGMVAGIATITKALKPSVKVYAAEP 211
DB 181 VVPVGGGMVAGIATITKALKPSVKVYAXXP 211
XX
XX RESULT 9
XX ABB63708
XX ID ABB63708 standard; Protein; 469 AA.
XX
XX AC ABB63708;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 17916.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.

```

[illegible]



```

XX Larrinua IM, Merlo DJ, Mourad GS, Paredy DR;
XX WPI; 1999-527375/44.
XX N-PSDB; AA211201.
XX
XX New nucleic acid encoding threonine dehydratase deaminase resistant
XX to feedback inhibition, useful as selection marker for cell
XX transformation and to impart herbicide resistance
XX
XX Claim 13; Page 111-114; 194pp; English.
XX
XX This sequence represents a mutant Arabidopsis thaliana threonine
XX dehydratase/deaminase (TD) protein of the invention. The protein is a
XX feedback insensitive mutant. The TD DNA sequence is used as molecular
XX marker (imparting resistance to toxic structural analogues of isoleucine)
XX for selecting transformed cells and to produce transformants with
XX increased levels of isoleucine (and thus better nutritional value) or of
XX intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for
XX synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also
XX TD-expressing plants permit use of the isoleucine structural analogues as
XX herbicides. The DNA sequences are alternatives for antibiotic resistance
XX markers (which are potentially harmful to the environment). Since no
XX human analog of TD exists (humans can not synthesize isoleucine), it
XX should be safe to use.
XX
XX Sequence 532 AA;
XX
XX Query Match 21.3%; Score 371; DB 20; Length 532;
XX Best Local Similarity 31.6%; Pred. No. 8e-29;
XX Matches 106; Conservative 61; Mismatches 134; Indels 34; Gaps 11;
XX
XX QY 4 QYCTSFADVEKAHINIODSTHLTPVLVTSSILNQIAGNLPFKCBLFOKTSFKIRGALNA 63
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 34 EYLTNLTSTKVYDIATESPLQAKLSKRL-----GYRMYLKREDLPQVFSFKLRGAYNM 88
XX
XX QY 64 IRLGIPDTPPEKPKAVVTHSGNHGQALTYAAKLEGIPAYIVVPOTAPNCKKLAIQAYGA 123
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 89 MWKLPAD--QLAKGVICSSAGNHAQGVALSASKLGCTAVIVMPVTPEIKWQAVENLGA 145
XX
XX QY 124 SIV-YCDPSPDESREKVTORTWQETEGI-LVHPNOEPAVIACQGTIALEVLNQV--PLVDA 179
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 146 TVVLFQGSYDQAQAAH--KIRAEPEGUTFPPDPDPIAGQGVFGMEITRQAKGPL-HA 202
XX
XX QY 180 LVYPVGGGGVMGVAITIKALKPSKVYVAAEPSNADDCYOSKLGELTNLHPPETI--- 236
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 203 IFPVGGGGGLIAGIAAYVKRVSPEVKIIGVEPADAN-----ANALSLHHGSERVILD 253
XX
XX QY 237 -----ADGVR--SSIGLNTWPIIRDLDVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVAL 290
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 254 QVGGFADGVAVKEVGEETFRISRNLMGVLVLTVDALCASIKDFEKNRILEPAGALAL 313
XX
XX QY 291 AAVLSOHQFTVSPVKNVCIVLSGGNVDLTSLNKKV 325
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 314 AGA-EAYCKYGYGLKDVNVVAITSGANNFDFKLRIV 347
XX
XX RESULT 14
XX AAY05707
XX ID AAY05707 standard; Protein: 532 AA.
XX AC
XX AC AAY05707;
XX
XX DT 19-JUL-1999 (first entry)
XX
XX Feedback insensitive mature threonine dehydratase/deaminase.
XX
XX Threonine dehydratase/deaminase: omr1 gene; feedback inhibition;
XX transgenic plant; selectable marker; isoleucine; mutant.
XX
XX Arabidopsis thaliana.
XX Synthetic.
XX

```

FH Key Location/Qualifiers  
 FT Misc-difference 439 /note= "Arg in wild-type enzyme"  
 FT Misc-difference 484 /note= "Arg in wild-type enzyme"  
 FT Region 427..444  
 FT /note= "regulatory region R4"  
 FT Region 476..494  
 FT /note= "regulatory region R6"  
 XX WO9902656-A1.  
 PN 21-JAN-1999.  
 XX 10-JUL-1998; 98WO-US14362.  
 XX 17-FEB-1998; 98US-0074875.  
 PR 10-JUL-1997; 97US-0052096.  
 XX (PURD ) PURDUE RES FOUND.  
 PA Mourad GS;  
 XX WPI; 1999-120860/10.  
 DR N-PSDB; AAX25336.  
 XX New sequences encode mutant threonine dehydratase/deaminase - which  
 PT is insensitive to feedback inhibition, useful as a selective marker  
 PT to produce transformed cells resistant to toxic isoleucine analogues  
 XX Disclosure: Page 65-68; 120pp; English.  
 PS This sequence represents an Arabidopsis thaliana mutant mature  
 CC threonine dehydratase/deaminase (TD) which, unlike the wild-type  
 CC enzyme (see AAY05702), is insensitive to feedback inhibition by  
 CC isoleucine. It lacks the chloroplast transit peptide of the  
 CC mutant TD precursor, and is encoded by a polynucleotide (see  
 CC AAX25336) that is one of 9 claimed polynucleotides (see AAX25332-40),  
 CC originally isolated and cloned from A. thaliana mutant line GM1lb  
 CC (omrl/omrl), which encode a feedback insensitive TD. These  
 CC polynucleotides can be used to transform a wide variety of plants,  
 CC fungi, bacteria and yeast. These polynucleotides differ from the  
 CC wild-type only by 2 point mutations, which result in R499C and  
 CC R554H amino acid substitutions (numbering according to wild-type  
 CC TD) in the R4 and R6 regulatory regions. These forms of TD are not  
 CC only insensitive to feedback inhibition by isoleucine, but are also  
 CC insensitive to structural analogues of isoleucine that are toxic to  
 CC plants and microorganisms which synthesize only wild-type TD.  
 CC Nucleotide sequences encoding mutated forms of TD can therefore be  
 CC used to create cells that are insensitive to compounds normally  
 CC toxic to cells expressing only wild-type TD enzymes, and thus may  
 CC be used to provide a biochemical selectable marker. Transformants  
 CC harboring a nucleotide sequence comprising a promoter operably  
 CC linked to a mutated TD sequence demonstrate increased levels of  
 CC isoleucine production, and thus provide an improved nutrient source.  
 XX Sequence 532 AA;  
 SQ  
 Query Match 21.3%; Score 371; DB 20; Length 532;  
 Best Local Similarity 31.6%; Pred. No. 8e-29;  
 Matches 106; Conservative 61; Mismatches 134; Indels 34; Gaps 11;  
 QY 4 CYCTSFADVERAHNIODSIHLTPVLTSSILNQIAGRNLFKCELFQKSGFKIRGALNA 63  
 DB 34 EYLNILSTKYVDIAESPLQAKLSKRL-----GVRWYLRDQLQVFSFKRGAVNM 88  
 QY 64 IRLGIPDTPPEKPAVYTHSSGNHQALTYAAKLEGIPAYIVVPQTAPNCKKLAIQAYGA 123  
 DB 89 MVKLPAD---QLAKGVICSSAGNHAQGVALSASLKGCTAVIVMPVTPPEIKKQAVENLGA 145  
 QY 124 SIV-YCDPSDSREKVTQRIHQETEGI-LVHPNDEPAVIAQGGTIALEVLNQV--PLVDA 179  
 DB 146 TVVLFGDSYDQAQAAH--KIRAEEGELTFIPFDHPDVIAGQGTVGMEITRQAKGPL-HA 202

QY 180 LVVPVGGGMVAGIAITIKALKPSKVVYAAEPSNADDCYQSKLKGELTPNLHPPETI--- 236  
 DB 203 IFVPVGGGLIAGIAAYVYKRVSPVKIIGVEPADAN-----AMALSUHGGERVILD 253  
 QY 237 -----ADGVK-SSIGLNTWPIIRDLDVDFVVTDEIKYATQLVWGRMKLLIETAGVAL 290  
 DB 254 QVGGFADGVAVKEVGEETFRISRNLMGVLVTRDAICASIKDMFEERKNIPEAGALAL 313  
 QY 291 AAVLSQHFQTVSPVKVNCIVLSGGNVDLTSLNWV 325  
 DB 314 AGA-EAYCKYGLKDVNVVAITSGANMFDKLRIV 347  
 RESULT 15  
 AAY32942  
 ID AAY32942 standard; Protein: 539 AA.  
 XX  
 AC AAY32942;  
 XX 09-NOV-1999 (first entry)  
 DT Mutant threonine dehydratase/deaminase protein sequence.  
 DE  
 XX Threonine dehydratase/deaminase: TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; muten.  
 XX Arabidopsis thaliana.  
 OS Synthetic.  
 XX WO9941395-A1.  
 PN 19-AUG-1999.  
 PD 08-JAN-1999; 99WO-US00560.  
 PF 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX (DOMC ) DOW AGROSCIENCES LLC.  
 XX (PURD ) PURDUE RES FOUND.  
 PI Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 XX WPI; 1999-527375/44.  
 DR N-PSDB; AAZ11200.  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX Claim 13; Page 106-109; 194pp; English.  
 PS This sequence represents a mutant Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
 CC marker (imparting resistance to toxic structural analogues of isoleucine)  
 CC for selecting transformed cells and to produce transformants with  
 CC increased levels of isoleucine (and thus better nutritional value) or of  
 CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for  
 CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
 CC TD-expressing plants permit use of the isoleucine structural analogues as  
 CC herbicides. The DNA sequences are alternatives for antibiotic resistance  
 CC markers (which are potentially harmful to the environment). Since no  
 CC human analog of TD exists (humans can not synthesize isoleucine), it  
 CC should be safe to use.  
 XX Sequence 539 AA;  
 SQ Query Match 21.3%; Score 371; DB 20; Length 539;  
 Best Local Similarity 31.6%; Pred. No. 8.1e-29;

Matches	106;	Conservative	61;	Mismatches	134;	Indels	34;	Gaps	11;
QY	4	QYCISFADVEKAHINIQDSIHITPVLTSILNQIAGRNLFKCELEFKTGSFKIRGALNA	63						
Ddb	41	EYLTNLSPKVYDIAIESPLQAKLKSRL----	GVRYLKREDLPVFSEKLRCAYNM	95					
QY	64	IGLIPDPTEEPKPAAVVTHSSNGHQALYYAALKLEGIPAYIVVPQTAPNCKKLAIOAYGA	123						
Ddb	96	MYKLPA-----QLAKGYICSAGNHQAOGVALSASKLGCTAVIVMPVTTPETIKWAOVENLGA	152						
QY	124	STV-YCDPSDESREKVQRIMQETEGI-LVHPNQEPAVTAGOCTIALEVNLQV--	PLVDA	179					
Ddb	153	TVLFGSDSYDQAQHA--KIRAEEEGLTFIPPDDHDPDTAQOCTVGMEITRQAKGPL-HA	209						
QY	180	LVPVPGGGGWAGIAITIKALXPSVKVYAAPSNAODCYQSCLKGLTLPMLPHPETI---	236						
Ddb	210	IFVPVGGGLIAGIAAYVRVSPEVKIIIGVEPADAN-----AMALSHHIGERVILD	260						
QY	237	-----ADGVK-SSIGLNTWPIIRDLDVDDFTVTTEDEIKYATQLVWGRMKLLIEPTAGVAL	290						
Ddb	261	QVGGFDAGVAVRVEGETFTRSNLMGDVVLVTRDAICASIKDMFEKRNILEPAGALAL	320						
QY	291	RAVLSQLHFQTVSPVKNVNCIVLSGGNVDTSLNWV	325						
Ddb	321	AGA-EAYCXYGIKDVNYVAITSGANNFEDKLRIV	354						

Search completed: June 24, 2003, 06:18:37  
Job time : 47.9323 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 06:23:16 ; Search time 285 Seconds  
(without alignments)  
8083.489 Million cell updates/sec

Title: US-09-889-609B-9  
Perfect score: 1023  
Sequence: 1 atgtgtctcagttatgcattt.....atcagtcgtttctgttttaa 1023

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues 4370478  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :				N_Geneseq_101002:*	
	1:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*			
	2:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*			
	3:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*			
	4:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*			
	5:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*			
	6:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*			
	7:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*			
	8:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*			
	9:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*			
	10:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*			
	11:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*			
	12:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*			
	13:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*			
	14:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*			
	15:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*			
	16:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*			
	17:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*			
	18:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*			
	19:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*			
	20:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*			
	21:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*			
	22:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*			
	23:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*			
	24:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1023	100.0	1023	21	AAA59299
2	1023	100.0	1023	22	AA515217
3	1023	100.0	1023	22	AAAD1461
4	1023	100.0	2477	22	AAH16282
5	1021.4	99.8	1336	22	AAI70575
6	1021.4	99.8	2674	22	AAI65020
7	816.6	79.8	1672	21	AAA59300
8	807.8	79.0	1018	21	AAA59294
9	729.2	71.3	848	22	AAH06600

10	645.6	63.1	731	24	ABL90122	Human polynucleoti
11	535	52.3	608	21	AAA59295	N-terminal sequenc
12	337	32.9	509	21	AAA59296	C-terminal sequenc
13	221	21.6	20892	22	ABA15709	Human nervous syst
14	200.2	19.6	861	21	AAAC10869	Human secreted pro
15	139.8	13.7	31096	18	AAV74370	Staphylococcus aur
16	112.4	11.0	583	22	AAH98041	Murine 7-transmemb
17	104	10.2	1447	23	ABL13783	Drosophila melanog
18	99.4	9.7	1189	23	ABU18395	Drosophila melanog
19	96.6	9.4	1509	20	AAZ11199	Mutant threonine d
20	96.6	9.4	1509	20	AAZ25334	Feedback insensiti
21	96.6	9.4	1599	20	AAZ11201	Mutant threonine d
22	96.6	9.4	1599	20	AAZ25336	Feedback insensiti
23	96.6	9.4	1620	20	AAZ11200	Mutant threonine d
24	96.6	9.4	1620	20	AAZ25335	Feedback insensiti
25	96.6	9.4	1638	20	AAZ11205	Mutant threonine d
26	96.6	9.4	1638	20	AAZ25340	Feedback insensiti
27	96.6	9.4	1779	20	AAZ11197	Mutant threonine d
28	96.6	9.4	1779	20	AAZ25331	Wild type threonin
29	96.6	9.4	1779	20	AAZ25332	Arabidopsis wild-t
30	96.6	9.4	1830	20	AAZ11198	Feedback insensiti
31	96.6	9.4	1830	20	AAZ25333	Mutant threonine d
32	96.6	9.4	2277	20	AAZ11212	Feedback insensiti
33	96.6	9.4	2277	20	AAZ11212	Mutant threonine d
34	94	9.2	542	22	AAI83752	Human polynucleoti
35	93.4	9.1	2235	20	AAZ11208	Mutant threonine d
36	93.4	9.1	2241	20	AAZ11206	Mutant threonine d
37	89.4	8.7	1830121	17	AAAT42063	Haemophilus influe
38	86.6	8.5	3550	23	AA588485	DNA encoding novel
39	84.8	8.3	1478	20	AA507180	Corn threonine dea
40	77.4	7.6	105184	24	ABK24122	Bacterial artifici
41	75.4	7.4	1866	23	AA582540	DNA encoding novel
42	75.4	7.4	1866	23	AA594486	DNA encoding novel
43	72.4	7.1	922	18	AAAT84000	DNA encoding a thr
44	71.6	7.0	3402	23	ABL18394	Drosophila melanog
45	71.6	7.0	4179	23	ABL13782	Drosophila melanog

ALIGNMENTS

RESULT 1  
AAA59299  
ID AAA59299 standard; DNA; 1023 BP.  
AC  
XX  
AA59299;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE DNA encoding a human serine racemase polypeptide.  
XX  
KW Serine racemase; N-methyl-D-aspartate receptor; neural death;  
KW neural dysfunction; NMDA receptor; Parkinson's disease;  
KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1023  
FT FT /\*tag= a  
FT FT /product= "serine racemase"  
XX  
PN WO200043526-A1.  
XX  
PD 27-JUL-2000.  
XX  
PF 18-JAN-2000; 2000WO-US00938.  
XX  
PR 19-JAN-1999; 99US-0116333.  
PR 21-JUL-1999; 99US-0144839.  
PR 28-JUL-1999; 99US-0145953.  
XX  
PA (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

XX Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;  
PI Ferris CD;  
XX WPI: 2000-482915/42.  
DR P-PSDB: AAB07734.  
XX Mammalian serine racemase preparations, used\* to identify modulators  
PT which can be used to treat diseases associated with  
PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease  
PT  
XX  
PS Claim 19; Page 49-50; 54pp; English.  
XX  
CC The present sequence encodes a mammalian serine racemase, which has  
CC a specific activity of at least 0.003 micromole L-serine/mg/hour.  
CC The enzyme catalyses the direct racemisation of L-serine to D-serine.  
CC D-serine appears to be an endogenous ligand of N-methyl-D-aspartate  
CC (NMDA) receptors. The mammalian serine racemases can be used to identify  
CC modulators, which can be used in the treatment of acute or chronic  
CC neuronal death or dysfunction mediated by overactivation of N-methyl-D-  
CC aspartate (NMDA) receptors. Overactivation of the receptors is  
CC associated with Parkinson's disease, Huntington's disease, motor neurone  
CC disease and Alzheimer's disease.  
XX  
S0 Sequence 1023 BP; 290 A; 220 C; 243 G; 270 T; 0 other;

Query Match 100.0%; Score 1023; DB 21; Length 1023;  
Best Local Similarity 100.0%; Pred. No. 2.3e-310;  
Matches 1023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGCTCAGTATTCATCTCTCTTCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
DB 1 ATGTGCTCAGTATTCATCTCTCTTCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
QY 61 GATTCATCACCTCACACAGTCTAACAGCTCCATTTTGAATCAACTAACAGGGCGC 120  
DB 61 GATTCATCACCTCACACAGTCTAACAGCTCCATTTTGAATCAACTAACAGGGCGC 120  
QY 121 AATCTTTCTTCAATGTGAACCTCTCCAGAAACAGGATCTTTTAAAGATTCGTGGTCT 180  
DB 121 AATCTTTCTTCAATGTGAACCTCTCCAGAAACAGGATCTTTTAAAGATTCGTGGTCT 180  
QY 181 CTCATGCTCAGAGCTGTTGGTCTCTGATGCTTTAGAAAAGAGCCGAAAGCTGTGTT 240  
DB 181 CTCATGCTCAGAGCTGTTGGTCTCTGATGCTTTAGAAAAGAGCCGAAAGCTGTGTT 240  
QY 241 ACTCACAGCAGTGAACCATGCGCCAGGCTCTCACCTATGCTGCCAAATTTGAAGGAAT 300  
DB 241 ACTCACAGCAGTGAACCATGCGCCAGGCTCTCACCTATGCTGCCAAATTTGAAGGAAT 300  
QY 301 CTTGCTTATATTTGTTGCTCCAGACAGCTCCAGACTGTAAAAAATTTGCAATACAGCC 360  
DB 301 CTTGCTTATATTTGTTGCTCCAGACAGCTCCAGACTGTAAAAAATTTGCAATACAGCC 360  
QY 361 TAGCGAGCGTCAATTTGATATCTGTGAACCTAGTGTGATGATGATGATGATGATGATG 420  
DB 361 TAGCGAGCGTCAATTTGATATCTGTGAACCTAGTGTGATGATGATGATGATGATGATG 420  
QY 421 AGAGTTACAGAAAGAACAGAGGATCATGTTGATGATGATGATGATGATGATGATGATG 480  
DB 421 AGAGTTACAGAAAGAACAGAGGATCATGTTGATGATGATGATGATGATGATGATGATG 480  
QY 481 GCTGGACAGGACAAATTTGCTGGAAGTCTGCAACAGGTTCTTTTGGTGGATGCACTG 540  
DB 481 GCTGGACAGGACAAATTTGCTGGAAGTCTGCAACAGGTTCTTTTGGTGGATGCACTG 540  
QY 541 GTGGTACCTTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTCTG 600  
DB 541 GTGGTACCTTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTCTG 600  
QY 601 AAACCTAGTGTGAGGTATATGCTGTAACCTCAATGTCAGATGCTCTACAGTCC 660  
DB 601 AAACCTAGTGTGAGGTATATGCTGTAACCTCAATGTCAGATGCTCTACAGTCC 660

DB 601 AAACCTAGTGTGAGGTATATGCTGTAACCTCAATGTCAGATGCTCTACAGTCC 660  
QY 661 AAGCTGAAGGGAAACTGATCCCAATCTTTATCTCCAGAAAACCATAGCAGATGGTCTC 720  
DB 661 AAGCTGAAGGGAAACTGATCCCAATCTTTATCTCCAGAAAACCATAGCAGATGGTCTC 720  
QY 721 AAATCCAGCATTTGGCTTTGAACACCTGGCTTATTTATCAGGACCTTTGTGGATGATATCT 780  
DB 721 AAATCCAGCATTTGGCTTTGAACACCTGGCTTATTTATCAGGACCTTTGTGGATGATATCT 780  
QY 781 ACTGTCACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGGATGAAACTA 840  
DB 781 ACTGTCACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGGATGAAACTA 840  
QY 841 CTCATTTCAACCTACAGCTGGTGTGGAGTGGCTGCTGCTGCTCAACATTTTCAAACT 900  
DB 841 CTCATTTCAACCTACAGCTGGTGTGGAGTGGCTGCTGCTGCTCAACATTTTCAAACT 900  
QY 901 GTTTCCCCAGAGTAAAGAACATTTGTTGCTGCTCAGTGGTGGAAATGTAGAGTTAACC 960  
DB 901 GTTTCCCCAGAGTAAAGAACATTTGTTGCTGCTCAGTGGTGGAAATGTAGAGTTAACC 960  
QY 961 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTTCTGTT 1020  
DB 961 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTTCTGTT 1020  
QY 1021 TAA 1023  
DB 1021 TAA 1023

RESULT 2  
AAS15217  
ID AAS15217 standard; cDNA; 1023 BP.  
XX AAS15217;  
AC AAS15217;  
XX 16-JAN-2002 (first entry)  
XX Human cDNA encoding Serine Racemase.  
XX Human; ss; serine racemase; Parkinson's disease; Huntington's disease;  
KW anxiety; glaucoma; stroke; hyperalgesia; pain; spinocerebellar ataxia;  
KW schizophrenia; transgenic animal; Chromosome 17p13.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH CDS 1..1023  
FT /\*tag= a  
FT /product= "Serine racemase"  
XX WO200175144-A1.  
XX 11-OCT-2001.  
XX 02-APR-2001; 2001WO-US10662.  
XX 04-APR-2000; 2000US-194451P.  
XX (MERI ) MERCK & CO INC.  
XX Connolly T, Liu Y, Xia M;  
XX WPI: 2001-656991/75.  
XX P-PSDB: AAU09124.  
XX New recombinant serine racemase polypeptide, useful in assays for  
PT identifying compounds that alter enzyme activity (e.g. including  
PT compounds that inhibit or stimulate enzyme activity) or in generating  
PT antibodies against the protein  
XX  
PS Claim 1; Page 9; 43pp; English.



DR P-PSDB; AAE08342.

XX Novel polypeptide of the human pyridoxal phosphate dependent family  
 PT useful in screening and detection assays and for treatment, e.g. of  
 PT epilepsy and Alzheimer's

XX Claim 1; Fig 1; 12lpp; English.

PS The present sequence is a cDNA encoding human pyridoxal phosphate  
 CC dependent enzyme 22406 which is a serine racemase. Human 22406 gene is  
 CC located on chromosome 17 between D17S849 and D17S996. The protein 22406  
 CC is a modulator of D-serine. D serine has been shown to modify  
 CC behavioural changes associated with learning, memory and convulsions.  
 CC Human 22046 and compounds that modulate the expression or activity are  
 CC used to treat or diagnose neurodegenerative disorders including  
 CC Alzheimer's disease, schizophrenia as well as quell anxiety and  
 CC epilepsy and prevent damage from stroke as well as cardiac (heart  
 CC failure, rheumatic heart failure) and circulatory disorders, liver  
 CC disorders (hepatic injury, jaundice), lung disorders, prostate  
 CC disorders (benign enlargement, nodular hyperplasia), colon disorders  
 CC (atresia, stenosis), skeletal muscle disorders (tumours-rhabdomyosarcoma)  
 CC and dermal fibroblast disorders. Human 22406 cDNA is also useful in gene  
 CC therapy.

SQ Sequence 1770 BP; 499 A; 379 C; 410 G; 482 T; 0 other;

Query Match 100.0%; Score 1023; DB 22; Length 1770;

Best Local Similarity 100.0%; Pred. No. 3.1e-310;  
 Matches 1023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTGCTCAGTATGTCATCTCCTTCTGATGTTGAAAAGCTTCATATCAACATTCGA 60  
 Db 69 ATGTTGCTCAGTATGTCATCTCCTTCTGATGTTGAAAAGCTTCATATCAACATTCGA 128  
 QY 61 GATTCTATCCACTCACACAGCTGCTACACAGTCCATTTTGAATCACTACAGGGCGC 120  
 Db 129 GATTCTATCCACTCACACAGCTGCTACACAGTCCATTTTGAATCACTACAGGGCGC 188  
 QY 121 AATCTTTTCTCAATGTGAATCTTCCAGAAAACAGGATCTTTTAAAGATCTGGTGCT 180  
 Db 189 AATCTTTTCTCAATGTGAATCTTCCAGAAAACAGGATCTTTTAAAGATCTGGTGCT 248  
 QY 181 CTCATGCGCTCAGAAAGCTTGGTCTCTGATGCTTTAGAAAAGAAAGCCGAAAGCTTGT 240  
 Db 249 CTCATGCGCTCAGAAAGCTTGGTCTCTGATGCTTTAGAAAAGAAAGCCGAAAGCTTGT 308  
 QY 241 ACTCACAGCTGGAACCAATGGCCAGCTCTCACCTATGCTGCCAATTTGAAGGAAT 300  
 Db 309 ACTCACAGCTGGAACCAATGGCCAGCTCTCACCTATGCTGCCAATTTGAAGGAAT 368  
 QY 301 CTGCTTTATTTGTTGCTCCAGACAGCTCCAGACTTAAAAAACCTTGCATACAGCC 360  
 Db 369 CTGCTTTATTTGTTGCTCCAGACAGCTCCAGACTTAAAAAACCTTGCATACAGCC 428  
 QY 361 TACGAGCGCTCAATGTATGTAATCTGTAACCTAGTATGATGATCCAGAGAAAATTTGCAAAA 420  
 Db 429 TACGAGCGCTCAATGTATGTAATCTGTAACCTAGTATGATGATCCAGAGAAAATTTGCAAAA 488  
 QY 421 AGACTTACAGAGAAACAGAGGATCATGTCATCCACAGGAGCCCTGCAGTGATA 480  
 Db 489 AGACTTACAGAGAAACAGAGGATCATGTCATCCACAGGAGCCCTGCAGTGATA 548  
 QY 481 GCTGGACAAGGACAAATTTGCCCTGGAAGTGTGAACAGGTTTCCTTTGGTGGATGCACTG 540  
 Db 549 GCTGGACAAGGACAAATTTGCCCTGGAAGTGTGAACAGGTTTCCTTTGGTGGATGCACTG 608  
 QY 541 GTGGTACCTGTAGGTGGAGGAGGAATGCTTCTGCTGGAATGAGCAATTTACAGTTAAGGCTCTG 600  
 Db 609 GTGGTACCTGTAGGTGGAGGAGGAATGCTTCTGCTGGAATGAGCAATTTACAGTTAAGGCTCTG 668  
 QY 601 AAACCTAGTGTGAAGGTATATGCTGTGAACCCCTCAATGTCAGATGACTGTACCAAGTCC 660  
 Db 669 AAACCTAGTGTGAAGGTATATGCTGTGAACCCCTCAATGTCAGATGACTGTACCAAGTCC 728

QY 661 AAGCTGAAGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGC 720  
 Db 729 AAGCTGAAGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGC 788  
 QY 721 AAATCCAGCATTTGGCTTGAACACCTGGCCTATTATCAGGGACCTTTGTGGATGATATCTTC 780  
 Db 789 AAATCCAGCATTTGGCTTGAACACCTGGCCTATTATCAGGGACCTTTGTGGATGATATCTTC 848  
 QY 781 ACTGTCACAGAGGATGAATTAAGTGTCCACCCAGCTGGTGTGGAGAGGATGAACATA 840  
 Db 849 ACTGTCACAGAGGATGAATTAAGTGTCCACCCAGCTGGTGTGGAGAGGATGAACATA 908  
 QY 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTCTCTCAACATTTTCAAACT 900  
 Db 909 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTCTCTCAACATTTTCAAACT 968  
 QY 901 GTTTCCTCCAGAGTAAGAACATTTTGTATTTGCTCAGTGGTGGAAATGTAGACTTAACC 960  
 Db 969 GTTTCCTCCAGAGTAAGAACATTTTGTATTTGCTCAGTGGTGGAAATGTAGACTTAACC 1028  
 QY 961 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGTT 1020  
 Db 1029 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGTT 1088  
 QY 1021 TAA 1023  
 Db 1089 TAA 1091

## RESULT 4

AAH16282

ID AAH16282 standard; cDNA; 2477 BP.

AC AAH16282;

DT 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:15148.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

OS Homo sapiens.

PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0138776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX XX WPI; 2001-318749/34.

XX PR Primer sets for synthesizing polynucleotides, particularly the 5602

XX PT full-length cDNAs defined in the specification, and for the detection

XX PT and/or diagnosis of the abnormality of the proteins encoded by the

XX PT full-length cDNAs

XX PS Claim 8; SEQ ID 15148; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602

XX CC full-length cDNAs defined in the specification. Where a primer set

XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

SQ Sequence 2477 BP; 738 A; 535 C; 524 G; 680 T; 0 other;

Query Match 100.0%; Score 1023; DB 22; Length 2477;  
Best Local Similarity 100.0%; Pred. No. 3.8e-310;  
Matches 1023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGTCAGTATGTCATCTCTTGTCTGATGTTGAAAGCTCATATCAACATTCGA 60  
DB 69 ATGTGTCAGTATGTCATCTCTTGTCTGATGTTGAAAGCTCATATCAACATTCGA 128  
QY 61 GATTCATCCACCTCACACAGTGTCTAACAAGCTCCATTTGAATCAACTAACAGGGCG 120  
DB 129 GATTCATCCACCTCACACAGTGTCTAACAAGCTCCATTTGAATCAACTAACAGGGCG 188  
QY 121 AATCTTTTCTTCAAAATGTGAATCTTCCAGAAACAGGATCTTTAAAGATCGTGCT 180  
DB 189 AATCTTTTCTTCAAAATGTGAATCTTCCAGAAACAGGATCTTTAAAGATCGTGCT 248  
QY 181 CTCATGCGTCGAGAGCTGTTGCTCTGATGCTTTAGAAAGAGCCGAAAGCTGTTGT 240  
DB 249 CTCATGCGTCGAGAGCTGTTGCTCTGATGCTTTAGAAAGAGCCGAAAGCTGTTGT 308  
QY 241 ACTCACAGCTGGAACCATGGCCAGGCTCTCACCCTATGTCGCCAAATTTGAAGAATT 300  
DB 309 ACTCACAGCTGGAACCATGGCCAGGCTCTCACCCTATGTCGCCAAATTTGAAGAATT 368  
QY 301 CTTGCTTATATGTTGGTCCCGCAGACAGCTCCAGACTGTAAATACTTTGCAATACAGCC 360  
DB 369 CTTGCTTATATGTTGGTCCCGCAGACAGCTCCAGACTGTAAATACTTTGCAATACAGCC 428  
QY 361 TACGAGCGCTCAATTTGTTACTGTGAACCTAGTGTAGTCCAGAGAAATGTTGCAAAA 420  
DB 429 TACGAGCGCTCAATTTGTTACTGTGAACCTAGTGTAGTCCAGAGAAATGTTGCAAAA 488  
QY 421 AGAGTTACAGAGAAACAGAGGATCATGTTACATCCCAACAGGAGCCCTGCAGTGATA 480  
DB 489 AGAGTTACAGAGAAACAGAGGATCATGTTACATCCCAACAGGAGCCCTGCAGTGATA 548  
QY 481 GCTGCACAGGAGCAATTTGCCCTGGAAGTGTGTAACACAGGTTCCCTTTGTTGGATCACTG 540  
DB 549 GCTGCACAGGAGCAATTTGCCCTGGAAGTGTGTAACACAGGTTCCCTTTGTTGGATCACTG 608  
QY 541 GTGGTACCTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATTTACAGTTAAAGCTCTG 600  
DB 609 GTGGTACCTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATTTACAGTTAAAGCTCTG 668  
QY 601 AAACCTAGTGTGAAGGTATATGCTCTCAACCCCTCAATGAGATGACTGTACAGTCC 660  
DB 669 AAACCTAGTGTGAAGGTATATGCTCTCAACCCCTCAATGAGATGACTGTACAGTCC 728  
QY 661 AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCTCCAGAAACCATAGCAGATGCTGC 720  
DB 729 AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCTCCAGAAACCATAGCAGATGCTGC 788

QY 721 AAATCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGACCTTGTGGATGATATCTTC 780  
DB 789 AAATCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGACCTTGTGGATGATATCTTC 848  
QY 781 ACTGTCCACAGGATGAATTAAGTGTGCAACCCAGCTGGTGGGAGAGGATGAACCTA 840  
DB 849 ACTGTCCACAGGATGAATTAAGTGTGCAACCCAGCTGGTGGGAGAGGATGAACCTA 908  
QY 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTCTCTCAACATTTTCAAACT 900  
DB 909 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTCTCTCAACATTTTCAAACT 968  
QY 901 GTTTCCTCCAGAGTAAAGAACATTTGTATTTGTCTCAGTGGTGGGAAATAGACTTAACC 960  
DB 969 GTTTCCTCCAGAGTAAAGAACATTTGTATTTGTCTCAGTGGTGGGAAATAGACTTAACC 1028  
QY 961 TCCTCCATCAACTTGGTGAAGCAGCTGAAAGCCAGCTTCTATCAGTCTCTTCTGTT 1020  
DB 1029 TCCTCCATCAACTTGGTGAAGCAGCTGAAAGCCAGCTTCTATCAGTCTCTTCTGTT 1088  
QY 1021 TAA 1023  
DB 1089 TAA 1091  
RESULT 5  
AAI70575  
ID AAI70575 standard; cDNA; 1336 BP.  
XX AAI70575;  
AC AAI70575;  
XX 21-JAN-2002 (first entry)  
XX Human serine racemase cDNA.  
XX Serine racemase; human; D-serine; regulation;  
KW glutamate N-methyl-D-aspartate receptor; neurodegenerative disease;  
KW stroke; Alzheimer's disease; Parkinson's disease; pain; myopathy;  
KW nootropic; neuroprotective; cerebrotective; antiparkinsonian;  
KW analgesic; diagnosis; gene therapy; screening; ss.  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH 52..1074  
FT CDS /\*tag= a  
ET  
XX  
PN WO200173077-A2.  
XX  
PD 04-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-EP03668.  
XX  
XX 31-MAR-2000; 2000US-193748P.  
PR 03-APR-2000; 2000US-194249P.  
XX  
XX (FARB ) BAYER AG.  
PA  
XX  
XX Ramakrishnan S;  
XX  
XX WPI; 2001-648444/74.  
DR P-PSDB; AAM50262.  
XX  
XX Polynucleotide encoding serine racemase enzyme and the enzyme useful  
PT for screening reagents regulating the activity of the enzyme in a  
PT neuron disease caused by over- or under-activation of glutamate  
PT N-methyl-D-aspartate  
XX  
XX Claim 1; Fig 1; 66pp; English.  
XX  
XX The present sequence is that of cDNA encoding human serine racemase  
CC (see AAM50262). The polynucleotide can be used in the production

CC of recombinant serine racemase enzyme, or for the detection of  
CC serine racemase polynucleotides. Expression vectors and host cells  
CC are claimed. Serine racemase catalyses the conversion of L-serine  
CC to D-serine. Neuron damage following various nervous system diseases  
CC is often caused by activation of glutamate N-methyl-D-aspartate  
CC (NMDA) receptors in the brain. This receptor is activated by the  
CC binding of D-serine. Regulation of D-serine levels through  
CC regulation of serine racemase may therefore prevent or minimise  
CC neuron damage in neurogenic and myopathic disorders.  
CC neurodegenerative disorders such as Alzheimer's disease and  
CC Parkinson's disease, and disorders leading to peripheral and  
CC chronic pain. Serine racemase polypeptides and polynucleotides are  
CC used in claimed methods of screening for agents that modulate or  
CC decrease the activity of serine racemase. Also claimed is a  
CC pharmaceutical composition comprising either an expression vector  
CC that contains a serine racemase polynucleotide, or a reagent that  
CC modulates serine racemase enzyme activity. This is used to modulate  
CC serine racemase activity in a disease, particularly neuron damage or  
CC a neurodegenerative disease caused by the over- or under-activation  
CC of the glutamate NMDA receptor.  
CC  
XX  
SQ Sequence 1336 BP; 386 A; 268 C; 319 G; 362 T; 1 other;

Query Match 99.88; Score 1021.4; DB 22; Length 1336;  
Best Local Similarity 99.94; Pred. No. 8.5e-310;  
Matches 1022; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGTGCTCAGTATGTCATCTCTCTGCTGATGTTGCAAAAGCTCATATCAACATCGA 60  
DB |||||||  
QY 52 ATGTGCTCAGTATGTCATCTCTCTGCTGATGTTGAAAGCTCATATCAACATCGA 111  
DB |||||||  
QY 61 GATTCTATCCACCTCACACCAGTCTCAACAGCTCCATTTTGAATCAACTAAACAGGCGC 120  
DB |||||||  
QY 112 GATTCTATCCACCTCACACCAGTCTCAACAGCTCCATTTTGAATCAACTAAACAGGCGC 171  
DB |||||||  
QY 121 AATCTTTTCTCAAAATGTGAACCTCTCCAGAAACAGGATCTTTTAAAGATTCGTGTCT 180  
DB |||||||  
QY 172 AATCTTTTCTCAAAATGTGAACCTCTCCAGAAACAGGATCTTTTAAAGATTCGTGTCT 231  
DB |||||||  
QY 181 CTCAATGCCCTCAGAGCTTGGTCTCTGATGCTTTAGAAAGGAGCGGAAGCTGTGTT 240  
DB |||||||  
QY 232 CTCAATGCCCTCAGAGCTTGGTCTCTGATGCTTTAGAAAGGAGCGGAAGCTGTGTT 291  
DB |||||||  
QY 241 ACTCACAGCAGTGGAAACCAATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGAAAT 300  
DB |||||||  
QY 292 ACTCACAGCAGTGGAAACCAATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGAAAT 351  
DB |||||||  
QY 301 CTTGCTTATATTTGTTGGTCCCGACAGAGCTCCAGACTGTAAAAAACCTGCAATACAGCC 360  
DB |||||||  
QY 352 CTTGCTTATATTTGTTGGTCCCGACAGAGCTCCAGACTGTAAAAAACCTGCAATACAGCC 411  
DB |||||||  
QY 361 TAGCGAGCGTCAATTTGTATCTGTGAACCTAGTGTAGTCCAGAGAAATGTTGCAAAA 420  
DB |||||||  
QY 412 TAGCGAGCGTCAATTTGTATCTGTGAACCTAGTGTAGTCCAGAGAAATGTTGCAAAA 471  
DB |||||||  
QY 421 AGAGTTACAGAGAAACAGAGGCATCATGGTATCATCCCAACAGGAGCTGCAGTGATA 480  
DB |||||||  
QY 472 AGAGTTACAGAGAAACAGAGGCATCATGGTATCATCCCAACAGGAGCTGCAGTGATA 531  
DB |||||||  
QY 481 GCTGGACAGGGACAATTTGCCCTGGAAGTCTGTAACCAAGGCTCTTTTGGTGGATGACATG 540  
DB |||||||  
QY 532 GCTGGACAGGGACAATTTGCCCTGGAAGTCTGTAACCAAGGCTCTTTTGGTGGATGACATG 591  
DB |||||||  
QY 541 GTGGTACCTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATTAACGTTAGGCTCTG 600  
DB |||||||  
QY 592 GTGGTACCTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATTAACGTTAGGCTCTG 651  
DB |||||||  
QY 601 AAACCTAGTGTGAAGTATATGCTGTAACCCCTCAAAATGCAGATGACTGCTACCAAGTCC 660  
DB |||||||  
QY 652 AAACCTAGTGTGAAGTATATGCTGTAACCCCTCAAAATGCAGATGACTGCTACCAAGTCC 711  
DB |||||||  
QY 661 AAGCTAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCAATAGCAATGCTGTC 720  
DB |||||||

DB 712 AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCTC 771  
QY 721 AATCCAGCATTTGGCTTTGACACCTGGCCTATTATCAGGACCTTTGGGATGATATCTTC 780  
DB |||||||  
DB 772 AATCCAGCATTTGGCTTTGACACCTGGCCTATTATCAGGACCTTTGGGATGATATCTTC 831  
QY 781 ACTGTCTACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGGAGAGGATGAAACTA 840  
DB |||||||  
DB 832 ACTGTCTACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGGAGAGGATGAAACTA 891  
QY 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTGCTCAACATTTTCAAACT 900  
DB |||||||  
DB 892 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTCAACATTTTCAAACT 951  
QY 901 GTTTCCTCCCAAGTAAGAACATTTGTTGCTGCTCAGTGGTGAATGTAGACTTAACC 960  
DB |||||||  
DB 952 GTTTCCTCCCAAGTAAGAACATTTGTTGCTCAGTGGTGAATGTAGACTTAACC 1011  
QY 961 TCCTCCATAACTTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTCTGTT 1020  
DB |||||||  
DB 1012 TCCTCCATAACTTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTCTGTT 1071  
QY 1021 TAA 1023  
DB |||||  
DB 1072 TAA 1074

RESULT 6  
AAI65020  
ID AAI65020 standard; cDNA; 2674 BP.  
XX AC AAI65020;  
XX XX  
DT 27-NOV-2001 (first entry)  
XX DE Serine/threonine dehydrase 37 coding sequence.  
XX KW Serine/threonine dehydrase 37; cancer; HIV infection; anti-HIV;  
KW cytosolic; ss.  
XX OS Unidentified.  
XX XX  
FH Key Location/Qualifiers  
FT CDS 262..1284  
FT /\*tag= a  
FT /product= "Serine/threonine dehydrase 37"  
XX  
PN CN1300824-A.  
XX XX  
PD 27-JUN-2001.  
XX XX  
PF 21-DEC-1999; 99CN-0125662.  
XX XX  
PR 21-DEC-1999; 99CN-0125662.  
XX {UYFU-} UNIV FUDAN.  
XX PA Mao Y, Xie Y;  
XX PI  
XX WPI; 2001-530468/59.  
DB P-PSDB; AAG78808.  
XX Polypeptide-serine/threonine dehydrase 37 and polynucleotide for coding  
PT this polypeptide -  
XX  
PS Claim 6; Page 25-26 (Disclosure); 34pp; Chinese.  
XX The present sequence is the coding sequence for serine/threonine  
CC dehydrase 37. The dehydrase and its coding sequence are useful for  
CC treating diseases e.g. cancer and HIV infection.  
XX  
SQ Sequence 2674 BP; 803 A; 581 C; 563 G; 727 T; 0 other;

Query Match		99.8%; Score 1021.4; DB 22; Length 2674;
Best Local Similarity		99.9%; Pred. No. 1.3e-309;
Matches 1022; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	1	ATGTGTGCTCAGTATTGCTATCTCTTTGCTGTGATGTTGAAAAAGCTCATATCAACATTCGA 60
DB	262	ATGTGTGCTCAGTATTGCTATCTCTTTGCTGTGATGTTGAAAAAGCTCATATCAACATTCGA 321
QY	61	GATTTCTATCCACTCACACAGTGTCTAACAAGCTCCATTTTGAATCAACTTAACAGGCGC 120
DB	322	GATTTCTATCCACTCACACAGTGTCTAACAAGCTCCATTTTGAATCAACTTAACAGGCGC 381
QY	121	AATCTTTTCTCAATGTGAATCTCTCCAGAAACAGGATCTTTTGAATCTGTGTGCT 180
DB	382	AATCTTTTCTCAATGTGAATCTCTCCAGAAACAGGATCTTTTGAATCTGTGTGCT 441
QY	181	CTCAATGCCGTGAGAAGCTTGGTCTGATGCTTTAGAAAGGAGCGCAAGCTGTGTT 240
DB	442	CTCAATGCCGTGAGAAGCTTGGTCTGATGCTTTAGAAAGGAGCGCGAAAGCTGTGTT 501
QY	241	ACTCACAGCAGTGGAAACCATGCGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAT 300
DB	502	ACTCACAGCAGTGGAAACCATGCGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAT 561
QY	301	CCTGCTTATATTGTGTGCCCCAGACAGCTCCAGACTGTAAAAACTTGCATACAGCC 360
DB	562	CCTGCTTATATTGTGTGCCCCAGACAGCTCCAGACTGTAAAAACTTGCATACAGCC 621
QY	361	TACGGAGCGTCAATTTGATCTGTGAACCTAGTGTGAGTCCAGAGAAATGTTGCAAAA 420
DB	622	TACGGAGCGTCAATTTGATCTGTGAACCTAGTGTGAGTCCAGAGAAATGTTGCAAAA 681
QY	421	AGAGTTACAGAAGAAACAGAAGCATCATGGTATACATCCCAACAGGAGCGCTGCAGTGATA 480
DB	682	AGAGTTACAGAAGAAACAGAAGCATCATGGTATACATCCCAACAGGAGCGCTGCAGTGATA 741
QY	481	GCTGGACAGGGACAAATGCTGCTGGAAGTGTCTGAACCAAGCTTCTTTGGTGGATGCACTG 540
DB	742	GCTGGACAGGGACAAATGCTGCTGGAAGTGTCTGAACCAAGCTTCTTTGGTGGATGCACTG 801
QY	541	GTGGTACCTGTAGTGTGAGGAGGAATGCTGTGGAATAGCAATTAAGTAAAGGCTCTG 600
DB	802	GTGGTACCTGTAGTGTGAGGAGGAATGCTGTGGAATAGCAATTAAGTAAAGGCTCTG 861
QY	601	AAACCTAGTGTGAAGTATATGCTGTGAACCTCAAATGCAATGCACTGCTACCACTCC 660
DB	862	AAACCTAGTGTGAAGTATATGCTGTGAACCTCAAATGCAATGCACTGCTACCACTCC 921
QY	661	AGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGC 720
DB	922	AGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGC 981
QY	721	AAATCCAGCATTTGGCTTGAACACTGGCTTATATCAGGAGCACTTCTGGATGATATCTTC 780
DB	982	AAATCCAGCATTTGGCTTGAACACTGGCTTATATCAGGAGCACTTCTGGATGATATCTTC 1041
QY	781	ACTGTCCACAGAGATGAAATTAAGTGTGAACCCAGCTGTGTGGAGAGGATGAACATA 840
DB	1042	ACTGTCCACAGAGATGAAATTAAGTGTGAACCCAGCTGTGTGGAGAGGATGAACATA 1101
QY	841	CTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT 900
DB	1102	CTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT 1161
QY	901	GTPTTCCCGAAGTGAAGAACATTTCTATTGTCTCAGTGGTGAATGTAGACTTAACC 960
DB	1162	GTPTTCCCGAAGTGAAGAACATTTCTATTGTCTCAGTGGTGAATGTAGACTTAACC 1221
QY	961	TCCTCCATACCTTGGGTGAAGCAGGCTGAAAGGCCAGCTTCTTATCAGTGTGTTTCTGTT 1020
DB	1222	TCCTCCATACCTTGGGTGAAGCAGGCTGAAAGGCCAGCTTCTTATCAGTGTGTTTCTGTT 1281
QY	1021	TAA 1023

DB	1282	TAA 1284
RESULT 7		
AAA59300		
ID	AAA59300	standard; DNA: 1672 BP.
XX	AAA59300;	
AC	AC	
DT	07-NOV-2000	(first entry)
XX	DNA encoding a murine serine racemase polypeptide.	
DE	Serine racemase; N-methyl-D-aspartate receptor; neural death;	
XX	neural dysfunction; NMDA receptor; Parkinson's disease;	
KW	Huntington's disease; motor neurone disease; Alzheimer's disease; ss.	
KW		
XX	Mus musculus.	
OS		
XX		
FH	Key	Location/Qualifiers
FT	CDS	219..1238
FT		/*tag= a
FT		/product= "serine racemase"
XX	WO200043526-A1.	
PN		
XX	27-JUL-2000.	
PD		
XX	18-JAN-2000; 2000WO-US000938.	
PF		
XX	19-JAN-1999; 99US-0116333.	
PR	21-JUL-1999; 99US-0144839.	
PR	28-JUL-1999; 99US-0145953.	
XX		
XX	(UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.	
PA		
XX	Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;	
PI	Ferris CD;	
PI		
XX	WPI: 2000-482915/42.	
DR	P-PSDB; AAB07731.	
DR		
XX	Mammalian serine racemase preparations, used to identify modulators	
PT	which can be used to treat diseases associated with	
PT	N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease	
PT		
XX	Disclosure; Fig 7A-B; 54pp; English.	
PS		
XX	The present sequence encodes a mammalian serine racemase, which has	
CC	a specific activity of at least 0.003 micromole L-serine/mg/hour.	
CC	The enzyme catalyses the direct racemisation of L-serine to D-serine.	
CC	D-serine appears to be an endogenous ligand of N-methyl-D-aspartate	
CC	(NMDA) receptors. The mammalian serine racemases can be used to identify	
CC	modulators, which can be used in the treatment of acute or chronic	
CC	neural death or dysfunction mediated by overactivation of N-methyl-D-	
CC	aspartate (NMDA) receptors. Overactivation of the receptors is	
CC	associated with Parkinson's disease, Huntington's disease, motor neurone	
CC	disease and Alzheimer's disease.	
XX		
SQ	Sequence 1672 BP; 491 A; 394 C; 379 G; 408 T; 0 other;	
Query Match		
Best Local Similarity 79.8%; Score 816.6; DB 21; Length 1672;		
Matches 901; Conservative 0; Mismatches 119; Indels 3; Gaps 1;		
QY	1	ATGTGTGCTCAGTATTGCTATCTCTTTGCTGTGATGTTGAAAAAGCTCATATCAACATTCGA 60
DB	219	ATGTGTGCTCAGTATTGCTATCTCTTTGCTGTGATGTTGAAAAAGCTCATATCAACATTCGA 278
QY	61	GATTTCTATCCACTCACACAGTGTCTAACAAGCTCCATTTTGAATCAACTTAACAGGCGC 120

Db 279 GACTCTATCCACCTCACCCAGTGGCTAACAGAGCTCCATTTTGAATCAATAGCAGGGCGC 338  
 Qy 121 AATCTTTTCTTCAAAATGAACTCTTCCAGAAAACAGGATCTTTTAAAGATTGCTGGTCT 180  
 Db 339 AATCTTTTCTTCAAAATGAACTCTTCCAGAAAACAGGATCTTTTAAAGATTGCTGGTCT 398  
 Qy 181 CTCATGCGCTGACAGCTTGGTCTTCCATGCTTTAGAAAAGAGCGGAAAGCTGTGTT 240  
 Db 399 CTTAATGCCATCAGAGGCTTAATTCCTGACACGCCAGAGAGAGCCCAAGCCGTAGTT 458  
 Qy 241 ACTCACAGCAGTGAACCAATGCGCAGGCTCTCTACCTATGCTGCCAATTCGAAGGATT 300  
 Db 459 ACTCACAGCAGCGGAACCAATGCGCAGGCTCTCTACCTATGCTGCCAATTCGAAGGATT 518  
 Qy 301 CCTGCTTATATGTGGTCCGCCAGACAGCTCCAGACTGTAAAAAATTTGCAATAACAGCC 360  
 Db 519 CCTGCTTATATGTGGTCCGCCAGACAGCTCCAGACTGTAAAAAATTTGCAATAACAGCC 578  
 Qy 361 TACGGAGGCTCAATTTGATATCTGTAACCTAGTAGTCCAGTCCAGAGAAAATTTGCAAAA 420  
 Db 579 TATGGAGCATCGATAGTATCTGTGACCCAAAGTGACGAGTCCAGAGAAAATTTGCAAAA 638  
 Qy 421 AGAGTTACAGAAAGAACAGAGGATCATGTTGATACATCCCAACAGGAGCTGCAAGTATA 480  
 Db 639 AGAATTATGCAAGAAACAGAGGATCATGTTGATACATCCCAACAGGAGCTGCAAGTATA 698  
 Qy 481 GCTGACAGGAGCAATTTGCCCTGGAAGTGTGTAACCAAGGTTCTTTGGTGGATGCACTG 540  
 Db 699 GCTGACAGGAGCAATTTGCCCTGGAAGTGTGTAACCAAGGTTCTTTGGTGGATGCACTG 758  
 Qy 541 GTGGTACTGTAGTGGAGGAGGAATGCTTGGTGGAAATAGCAATACAGTTAAAGGCTCTG 600  
 Db 759 GTGGTACTGTAGTGGAGGAGGAATGCTTGGTGGAAATAGCAATACAGTTAAAGGCTCTG 818  
 Qy 601 AAACCTAGTGTAGGATATATGCTGTAACCTCAACCTCAATGATGATGCTGCTACCACTCC 660  
 Db 819 AAACCTAGTGTAGGATATATGCTGTAACCTCAACCTCAATGATGATGCTGCTACCACTCC 878  
 Qy 661 AAGCTGAAGGGGAACTGATGCCCAATCTTTATCTCTCCAGAAACCATAGCAGATGCTGTC 720  
 Db 879 AAACCTGAAGGAGAACTGATGCCCAATCTTTATCTCTCCAGAAACCATAGCAGATGCTGTC 938  
 Qy 721 AAATCCAGCATTTGGCTGACACCTGGCCTTATATCAGGAGACCTTGTGGATGATATCTTC 780  
 Db 939 AAATCCAGCATTTGGCTGACACCTGGCCTTATATCAGGAGACCTTGTGGATGATATCTTC 998  
 Qy 781 ACTGTCACAGAGATGAAATTAAGTGTGCAACCCAGCTGCTGTGGAGAGGATGAAACTA 840  
 Db 999 ACTGTCACCAAGATGAAATCAAGTATGCAACCCAGCTGCTGTGGAGAGGATGAAACTG 1058  
 Qy 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGCTGTGCTCTCAACATTTTCAAACT 900  
 Db 1059 CTCATTGAGCGGACTGCTGGCGTGGCAGCTGGCTGCACTGCTCTCAGCATTTTCCAAACA 1118  
 Qy 901 GTTTCCCCAGAACTAAGACATTTTGTATGCTGCTCAGTGGTGAATAGTACTTAACC 960  
 Db 1119 GTCTCTCCAGAACTAAGACATTTTGTATGCTGCTCAGTGGTGAATAGTACTTAACC 1176  
 Qy 961 TCTCTCATTAATTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTTATCAGTCTGTTCTGTT 1020  
 Db 1177 -CTCTCTGACTGGGTGGGCGAGGCTGAAAGCCAGCTTCTTTATCAGTCTGTTCTGTT 1235  
 Qy 1021 TAA 1023  
 Db 1236 TAA 1238

RESULT 8

AAA59294

ID AAA59294 standard; DNA; 1018 BP.

XX

AC AAA59294;

XX

DT 07-NOV-2000 (first entry)  
 XX DNA encoding a murine serine racemase polypeptide.  
 DE  
 XX Serine racemase; N-methyl-D-aspartate receptor; neural death;  
 KW neural dysfunction; NMDA receptor; Parkinson's disease;  
 KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 DE 1..1018  
 FT CDS /\*tag= a  
 FT /transl\_except= (pos: 1012, aa: Ser)  
 FT /product= "serine racemase"  
 XX  
 PN WO200043526-A1.  
 XX  
 XX 27-JUL-2000.  
 XX  
 XX 18-JAN-2000; 2000WO-US00938.  
 XX  
 PR 19-JAN-1999; 99US-0116333.  
 PR 21-JUL-1999; 99US-0144839.  
 PR 28-JUL-1999; 99US-0145953.  
 XX  
 XX (UOJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 PA  
 XX Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;  
 PI Ferris CD;  
 XX  
 DR WPI: 2000-482915/42.  
 DR P-PSDB: AAB07731.  
 XX  
 XX Mammalian serine racemase preparations, used to identify modulators  
 PT which can be used to treat diseases associated with  
 PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease  
 PT  
 XX  
 XX Claim 16; Page 45-46; 54pp; English.  
 CC The present sequence encodes a mammalian serine racemase, which has  
 CC a specific activity of at least 0.003 micromole L-serine/mg/hour.  
 CC The enzyme catalyses the direct racemisation of L-serine to D-serine.  
 CC D-serine appears to be an endogenous ligand of N-methyl-D-aspartate  
 CC (NMDA) receptors. The mammalian serine racemases can be used to identify  
 CC modulators, which can be used in the treatment of acute or chronic  
 CC neural death or dysfunction mediated by overactivation of N-methyl-D-  
 CC aspartate (NMDA) receptors. Overactivation of the receptors is  
 CC associated with Parkinson's disease, Huntington's disease, motor neurone  
 CC disease and Alzheimer's disease.  
 XX  
 SQ Sequence 1018 BP; 290 A; 250 C; 243 G; 235 T; 0 other;  
 Query Match 79.0%; Score 807.8; DB 21; Length 1018;  
 Best Local Similarity 87.7%; Pred. No. 8.4e-243;  
 Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;  
 QY 1 ATGTGTGCTCAGTATTGTCATCTCTTTGCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
 DB 1 ATGTGTGCTCAGTATTGTCATCTCTTTGCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
 QY 61 GATTTCTATCCACTCACACAGTGTCAACAGCTCCATTTGTAATCAACTAACAGGGCGC 120  
 DB 61 GACTCTATCCACTCACACAGTGTCAACAGCTCCATTTGTAATCAACTAACAGGGCGC 120  
 QY 121 AATCTTTTCTTCAAAATGTAAGTCTTCCAGAAAACAGGATCTTTTAAAGATTGCTGGTCT 180  
 DB 121 AATCTTTTCTTCAAAATGTAAGTCTTCCAGAAAACAGGATCTTTTAAAGATTGCTGGTCT 180  
 QY 181 CTCATGCGCTGACAGCTTGGTCTGATGCTTTAGAAAGAGCGGAAAGCTGTGTT 240  
 DB 181 CTTAATGCCATCAGAGGCTTAATTCCTGACACGCCAGAGAGAGCCCAAGCCGTAGTT 240

```
Qy 241 ACTCAGCAGCTGGAACCAATGGCCAGGCTCTCACCTATCTGCTGCCAAATTTGGAAGGAATT 300
Db 241 ACTCAGCAGCAGGAAACCAATGGCCAGGCTCTCACCTATCTGCTGCCAAATTTGGAAGGAATT 300
Qy 301 CTTGCTTATATTGTTGGTGGCCAGACAGCTCCAGACTGTAAAAAATCTGCAATACAGGCC 360
Db 301 CTTGCTTATATTGTTGGTGGCCAGACAGCTCCAGACTGTAAAAAATCTGCAATACAGGCC 360
Qy 361 TAGCGAGCGTCAATTGTATATCTGTGAACCTAGTAGTCCAGAGAAAATGTTGCAAAA 420
Db 361 TAGCGAGCAGTATAGTATATCTGTGACCAAGTACAGAGTCCAGAGAAAATGTTGCAAAA 420
Qy 421 AGATTACAGAAACAGAGGATCATGGTACATCCCAACAGGAGCCTGCAGTGATA 480
Db 421 AGAATTATGAGAAACAGAGGATCATGGTACATCCCAACAGGAGCCTGCAGTGATA 480
Qy 481 GCTGGACACGGGCAATTGCTTGGGAAAGTCTGCAACAGGCTCTCTTTGGTGGATGCACTG 540
Db 481 GCTGGACAGGACAAATGCTTGGGAAAGTCTGCAACAGGCTCTCTTTGGTGGATGCACTG 540
Qy 541 GTGCTACCTAGTGGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTCTG 600
Db 541 GTGCTACCTAGTGGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTCTG 600
Qy 601 AAACCTAGTGAAGGTATATGCTGTGAACCTCAATCAATGCAGATGCTCTACCACTCC 660
Db 601 AAACCTAGTGAAGGTATATGCTGTGAACCTCAATCAATGCAGATGCTCTACCACTCT 660
Qy 661 AAGCTGAAGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCT 720
Db 661 AAGCTGAAGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCT 720
Qy 721 AAATCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGAGCCTTGTGGATGATCTTC 780
Db 721 AAATCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGAGCCTTGTGGATGATCTTC 780
Qy 781 ACTGTCACAGAGGATGAATTAAGTGTGCAACCCAGCTGTGTGGGAGAGAGTGAACATA 840
Db 781 ACTGTCACCAAGATGAATTAAGTGTGCAACCCAGCTGTGTGGGAGAGTGAACATA 840
Qy 841 CTATTGAACCTACAGCTGTGTGGAGTGGCTGTGCTGTCTGCTCAACATTTTCAAACT 900
Db 841 CTATTGAGCGGACTGTGGCGTGGCACTGGCTGTGCTGTCTGCTCAACATTTTCAAACT 900
Qy 901 GTTTCGCCAGAGTAAGAACATTTGTATTGTGCTCAGTGGTGGAAATGTAGACTTAACC 960
Db 901 GTCTCTCCAGAGTAAGAACAGTCTGCATTGTACTCAGTGGGGGAATGTAGACTTAA-- 958
Qy 961 TCCTCCATTAATCTGGGTGAAGCAGGCTGAAGGCCAGCTTCTTATCAGTCTGTTTCGT 1019
Db 959 -CCTCCCTGAATGGGTGGGCGAGGCTGAACGGCCAGCTCCTTACCAGAGCGTCTGTTT 1016

RESULT 9
AAH06600
ID AAH06600 standard; cDNA; 848 BP.
XX
AC AAH06600;
AC
DX 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:3435.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
```

XX  
PR  
PR  
PR  
PR  
PR  
XX  
PA  
XX  
PI  
PI  
XX  
DR  
XX  
PT  
PT  
PT  
XX  
PS  
XX

29-JUL-1999; 99JP-0248036.  
27-AUG-1999; 99JP-0300253.  
11-JAN-2000; 2000JP-0118776.  
02-MAY-2000; 2000JP-0183767.  
09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 1; SEQ ID 3435; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX  
SQ Sequence 848 BP; 236 A; 187 C; 214 G; 207 T; 4 other;

Query Match 71.3%; Score 729.2; DB 22; Length 848;  
Best Local Similarity 98.0%; Pred No. 3.6e-218;  
Matches 768; Conservative 0; Mismatches 12; Indels 4; Gaps 3;

Qy 1 ATGTGTCTCAGTATTGCTATCTCTCTTTGCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
Db 69 ATGTGTCTCAGTATTGCTATCTCTTTGCTGATGTTGAAAAGCTCATATCAACATTCGA 128

Qy 61 GATTCTATCCACCTCACACCAGTCTTAACAGTCCATTTTGAATCAACTAACAGGGCGC 120  
Db 129 GATTCTATCCACCTCACACCAGTCTTAACAGTCCATTTTGAATCAACTAACAGGGCGC 188

Qy 121 AATCTTTTCTTCAATGTGAACCTCTCCAGAAAACAGAGTCTTTTAAAGATTCTGGTGTCT 180  
Db 189 AATCTTTTCTTCAATGTGAACCTCTCCAGAAAACAGAGTCTTTTAAAGATTCTGGTGTCT 248

Qy 181 CTCAATGCCGTCAGAACTTTGGTTCCTGATGCTTTTAGAAAAGAGCGAAAGCTGTGTT 240  
Db 249 CTCAATGCCGTCAGAACTTTGGTTCCTGATGCTTTTAGAAAAGAGCGAAAGCTGTGTT 308

Qy 241 ACTCAGCAGTGAACCACTAGCCAGGCTCTCACTATGCTGCCAAATTTGGAAGGAATT 300  
Db 309 ACTCAGCAGTGAACCACTAGCCAGGCTCTCACTATGCTGCCAAATTTGGAAGGAATT 368

Qy 301 CCTGCTTATATTGTTGGTGGCCAGACAGCTCCAGACTGTAAAAAATTTGCAATACAGCC 360  
Db 369 CCTGCTTATATTGTTGGTGGCCAGACAGCTCCAGACTGTAAAAAATTTGCAATACAGCC 428

QY 361 TACGAGCGTCAATTGTTACTGTGAACCTAGTGTAGTCCAGAGAAATGTTCCAAA 420  
 Db |||||||  
 QY 429 TACGAGCGTCAATTGTTACTGTGAACCTAGTGTAGTCCAGAGAAATGTTCCAAA 488  
 Db |||||||  
 QY 421 AGAGTTACAGAAGAACAGAGGCATCATGTTACATCCCAACAGGAGCCTGCAGTGATA 480  
 Db |||||||  
 QY 489 AGAGTTACAGAAGAACAGAGGCATCATGTTACATCCCAACAGGAGCCTGCAGTGATA 548  
 Db |||||||  
 QY 481 GCTGACAGGAGCAATTCCTGGAAGTGTGAACCAAGGTTCTTTGCTGGATGCACTG 540  
 Db |||||||  
 QY 549 GCTGACAGGAGCAATTCCTGGAAGTGTGAACCAAGGTTCTTTGCTGGATGCACTG 608  
 Db |||||||  
 QY 541 GTGGTACCTGTAGTGGAGGAGGAATGCTTCTGGAATAGCAATPACAGTTAAAGGCTCTG 600  
 Db |||||||  
 QY 609 GTGGTACCTGTAGTGGAGGAGGAATGCTTCTGGAATAGCAATPACAGTTAAAGGCTCTG 668  
 Db |||||||  
 QY 601 AAACCTAGTGTGAAGGTATGCTGCAACCTCAATGAGATGAGTGTACAGTCC 660  
 Db |||||||  
 QY 669 AAACCTAGTGTGAAGGTATGCTGCAACCTCAATGAGATGAGTGTACAGTCC 727  
 Db |||||||  
 QY 661 AAGCTGAAGGGGAACTGATCCCAATCTTTATCTCCAGAAACATAGCAGATGGTGC 720  
 Db |||||||  
 QY 728 AAGCTGAAGGGGAACTGATCCCAATCTTTATCTCCAGAAACATAGCAGATGGTGC 785  
 Db |||||||  
 QY 721 AAATCCAGCATGGCTTGAACACCTGGCCTATTATCAGGGACCTTGTGGATGATATCTTC 780  
 Db |||||||  
 QY 786 AAATCCCA-CATTGGCTTGAACACCTGGNCCTATTATCANGGACCTTGGGATGATATCTTT 844  
 QY |||||  
 Db 845 ACTG 848

## RESULT 10.

ABL90122  
 ID ABL90122 standard; cDNA; 731 BP.  
 XX  
 AC ABL90122;  
 DT 24-MAY-2002 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 684.  
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein; gene; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200190304-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001WO-US16450.  
 XX  
 PR 19-MAY-2000; 2000US-205515P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Birse CE, Rosen CA;  
 XX  
 DR WPI; 2002-122018/16.  
 DR P-PSDB; ABB89713.  
 XX  
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive,  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders -  
 XX  
 PS Claim 4; SEQ ID NO 684; 2081pp + Sequence Listing; English.

XX

CC The invention relates to novel genes (ABL89449-ABL90853) and proteins  
 CC (ABB9040-ABB9044) useful for preventing, treating, or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.

CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 731 BP; 204 A; 158 C; 184 G; 178 T; 7 other;

Query Match 63.1%; Score 645.6; DB 24; Length 731;  
 Best Local Similarity 99.0%; Pred. No. 5.9e-192;

Matches 667; Conservative 1; Mismatches 4; Indels 2; Gaps 2;

QY 1 ATGTGTGCTCAGTATTGCTCTCTTCTGCTGATGTTGAAAGCTCATATCAACATTCGA 60  
 Db |||||||  
 QY 47 ATGTGTGCTCAGTATTGCTCTCTTCTGCTGATGTTGAAAGCTCATATCAACATTCGA 106  
 Db |||||||  
 QY 61 GATTCTATCCACCTCACACCAAGTCTTCCAGAAAGCTCAATTTGAATCAACTAACAGGCGC 120  
 Db |||||||  
 QY 107 GATTCTATCCACCTCACACCAAGTCTTCCAGAAAGCTCAATTTGAATCAACTAACAGGCGC 166  
 Db |||||||  
 QY 121 AATCTTTCTTCAATGTGAATCTTCCAGAAAGCTCAATTTGAATCAACTAACAGGCGC 180  
 Db |||||||  
 QY 167 AATCTTTCTTCAATGTGAATCTTCCAGAAAGCTCAATTTGAATCAACTAACAGGCGC 226  
 Db |||||||  
 QY 181 CTCAATGCCGTGCAAGCTTGGTCTTCTGATGTTGAAAGCTCAATTTGAATCAACTAACAGGCGC 240  
 Db |||||||  
 QY 227 CTCATGCCGTGCAAGCTTGGTCTTCTGATGTTGAAAGCTCAATTTGAATCAACTAACAGGCGC 286  
 Db |||||||  
 QY 241 ACTCACAGCAGTGGAAACCAATGGCCAGGCTCTCACTTATGCTGCCAAATTTGAAGGAAAT 300  
 Db |||||||  
 QY 287 ACTCACAGCAGTGGAAACCAATGGCCAGGCTCTCACTTATGCTGCCAAATTTGAAGGAAAT 346  
 Db |||||||  
 QY 301 CCTGCTTATATTGTTGGTGGCCCGACAGAGTCCAGAGCTGTAAGAACTTGCATACAGCC 360  
 Db |||||||  
 QY 347 CCTGCTTATATTGTTGGTGGCCCGACAGAGTCCAGAGCTGTAAGAACTTGCATACAGCC 406  
 Db |||||||  
 QY 361 TACGGAGCGTCAATTGTTATATCTGGAACCTAGTGTAGTCCAGAGAAATGTTGCAAAA 420  
 Db |||||||  
 QY 407 TACGGAGCGTCAATTGTTATATCTGGAACCTAGTGTAGTCCAGAGAAATGTTGCAAAA 466  
 Db |||||||  
 QY 421 AGAGTTACAGAAGAACAGAGGCATCATGTTACATCCCAACAGGAGCCTGCAGTGATA 480  
 Db |||||||  
 QY 467 AGAGTTACAGAAGAACAGAGGCATCATGTTACATCCCAACAGGAGCCTGCAGTGATA 526  
 Db |||||||  
 QY 481 GCTGGACAAGGAGCAATTCCTGGAAGTGTGTAACAGGTTCTTTGGTGGATGCACTG 540  
 Db |||||||  
 QY 527 GCTGGACAAGGAGCAATTCCTGGAAGTGTGTAACAGGTTCTTTGGTGGATGCACTG 586  
 Db |||||||  
 QY 541 GTGGTACCTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATTAAGGCTCTG 600  
 Db |||||||  
 QY 587 GTGGTACCTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATTAAGGCTCTG 646  
 Db |||||||  
 QY 601 AAACCTAGTGTGAAGGTATGCTGCTGAACCTCAATGAGATGAGTGTACAGTCC 660  
 Db |||||||  
 QY 647 AAACCTAGTGTGAAGGTATGCTGCTGTAACCTCAATGAGATGAGTGTACAGTCC 704  
 Db |||||||  
 QY 661 AAGCTGAAGGGGAA 674  
 Db |||||||  
 QY 705 AAGCTGAAGGGGAA 718  
 Db |||||||

RESULT 11	
AAA59295	
ID AAA59295 standard; DNA: 608 BP.	
XX AC AAA59295;	
XX 07-NOV-2000 (first entry)	
XX N-terminal sequence of human serine racemase DNA.	
XX Serine racemase; N-methyl-D-aspartate receptor; neural death;	
KW neural dysfunction; NMDA receptor; parkinson's disease;	
KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.	
XX Homo sapiens.	
XX WO200043526-A1.	
PN 27-JUL-2000.	
XX 18-JAN-2000; 2000WO-US00938.	
XX 19-JAN-1999; 99US-0116333.	
PR 21-JUL-1999; 99US-0144839.	
PR 28-JUL-1999; 99US-0145953.	
XX (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.	
XX Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;	
PI Ferris CD;	
XX WPI; 2000-482915/42.	
XX Mammalian serine racemase preparations, used to identify modulators	
PT which can be used to treat diseases associated with	
PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease	
XX Claim 17; Page 26; 54pp; English.	
XX The present sequence represents a fragment of a mammalian serine	
CC racemase gene. The racemase polypeptide has a specific activity of at	
CC least 0.003 micromole L-serine/mg/hour. The enzyme catalyses the	
CC direct racemisation of L-serine to D-serine. D-serine appears to be	
CC an endogenous ligand of N-methyl-D-aspartate (NMDA) receptors. The	
CC mammalian serine racemases can be used to identify modulators, which	
CC can be used in the treatment of acute or chronic neural death or	
CC dysfunction mediated by overactivation of N-methyl-D-aspartate (NMDA)	
CC receptors. Overactivation of the receptors is associated with	
CC Parkinson's disease, Huntington's disease, motor neurone disease and	
CC Alzheimer's disease.	
XX Sequence 608 BP; 174 A; 137 C; 148 G; 148 T; 1 other;	
SQ	
Query Match 52.3%; Score 535; DB 21; Length 608;	
Best Local Similarity 98.4%; Pred. No. 2.8e-157;	
Matches 571; Conservative 0; Mismatches 6; Indels 3; Gaps 3;	
QY 1 ATGTGTGTCAGTATTGCTATCTCTTGTGATGTTGAAAGAGCTATATCAACATTCGA 60	
DB 27 ATGTGTGTCAGTATTGCTATCTCTTGTGATGTTGAAAGAGCTATATCAACATTCGA 86	
QY 61 GATTCTATCCACTCACAGCTGTCTTGTGATGTTGAAAGAGCTATATCAACATTCGA 120	
DB 87 GATTCTATCCACTCACAGCTGTCTTGTGATGTTGAAAGAGCTATATCAACATTCGA 146	
QY 121 AATCTTTTCTTCAATGTGAAGTCTTCCAGAAACAGGATCTTTTAAGATTGCTGTGCT 180	
DB 147 AATCTTTTCTTCAATGTGAAGTCTTCCAGAAACAGGATCTTTTAAGATTGCTGTGCT 206	
QY 181 CTCATGCGCGTCAGAGCTTGGTTCCTGTATGCTTTTAGAAGAGGCCGGAAGCTGTGTT 240	

Db	207	CTCAATGCCGTCAGAAAGCTTGGTTCTGATGCTTTAGAAAGGAGCGGAAAGCTGTGTT	266
QY	241	ACTCACAGCAGTGGAAACCATGCCAGGCTCTCACCTATGCTCCAAATTTGGAGGAATT	300
Db	267	ACTCACAGCAGTGGAAACCATGCCAGGCTCTCACCTATGCTCCAAATTTGGAGGAATT	326
QY	301	CCTGCTTATATTCTGCTGCCGCCAGACAGCTCCAGACTGTAAAAAATTTGCAATACAGCC	360
Db	327	CCTGCTTATATTCTGCTGCCGCCAGACAGCTCCAGACTGTAAAAAATTTGCAATACAGCC	386
QY	361	TACGGAGCGTCAATTCTATCTGTGAACCTAGTGATG-AGTCCAGAGAAATTTGCCAAA	419
Db	387	TACGGAGCGTCAATTCTATCTGTGAACCTAGTGATG-AGTCCAGAGAAATTTGCCAAA	446
QY	420	AA-GAGTTACAGAAAGACAGAGGATCATGCTACATCCCAACAGGAGCCCTGCAGTGA	478
Db	447	AAGGAGTTACAGAAAGACAGAGGATCATGCTACATCCCAACAGGAGCCCTGCAGTGA	506
QY	479	TAGCTGCACAAGGAGCAATTGCCCTGGAAGTCTGAACAGGTTCCCTTTGGTGGATGCAC	538
Db	507	TAGCTGCACAAGGAGCAATTGCCCTGGAAGTCTGAACAGGTTCCCTTTGGTGGATGCAC	566
QY	539	TGGTGTACTCTGTAGTGG-AGGAGGAATGCTTGTCTGGAA 577	
Db	567	TGGTGTGNCCTGTAGTGGAGGAGGAATGCTTGTCTGGGGA 606	

RESULT 12

AAA59296

ID AAA59296 standard; DNA: 509 BP.

XX AC AAA59296;

XX 07-NOV-2000 (first entry)

XX C-terminal sequence of human serine racemase DNA.

XX Serine racemase; N-methyl-D-aspartate receptor; neural death;

KW neural dysfunction; NMDA receptor; parkinson's disease;

KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.

XX Homo sapiens.

XX WO200043526-A1.

XX 27-JUL-2000.

XX 18-JAN-2000; 2000WO-US00938.

XX 19-JAN-1999; 99US-0116333.

PR 21-JUL-1999; 99US-0144839.

PR 28-JUL-1999; 99US-0145953.

XX (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

XX Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;

PI Ferris CD;

XX WPI; 2000-482915/42.

XX Mammalian serine racemase preparations, used to identify modulators

PT which can be used to treat diseases associated with

PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease

XX Claim 18; Page 27; 54pp; English.

XX The present sequence represents a fragment of a mammalian serine

CC racemase gene. The racemase polypeptide has a specific activity of at

CC least 0.003 micromole L-serine/mg/hour. The enzyme catalyses the

CC direct racemisation of L-serine to D-serine. The enzyme catalyses the

CC direct racemisation of L-serine to D-serine. D-serine appears to be

CC an endogenous ligand of N-methyl-D-aspartate (NMDA) receptors. The

CC mammalian serine racemases can be used to identify modulators, which

CC can be used in the treatment of acute or chronic neural death or

CC dysfunction mediated by overactivation of N-methyl-D-aspartate (NMDA)

CC receptors. Overactivation of the receptors is associated with

CC Parkinson's disease, Huntington's disease, motor neurone disease and

CC Alzheimer's disease.

XX Sequence 608 BP; 174 A; 137 C; 148 G; 148 T; 1 other;

SQ

CC mammalian serine racemases can be used to identify modulators, which  
 CC can be used in the treatment of acute or chronic neural death or  
 CC dysfunction mediated by overactivation of N-methyl-D-aspartate (NMDA)  
 CC receptors. Overactivation of the receptors is associated with  
 CC Parkinson's disease, Huntington's disease, motor neurone disease and  
 CC Alzheimer's disease.  
 XX  
 SQ Sequence 509 BP; 139 A; 90 C; 115 G; 164 T; 1 other;

Query Match 32.98; Score 337; DB 21; Length 509;  
 Best Local Similarity 99.78; Pred. No. 3.7e-95;  
 Matches 348; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 Oy 676 CTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGCAAAATCCAGCATGGC 735  
 Db 1 CTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGCAAAATCCAGCATGGC 60  
 Oy 736 TTGAA-CACCTGGCCTATTATCAAGGACCTTTGGATGATATCTTCACTGTCAAGAGGA 794  
 Db 61 TTGAANACCTGGCCTATTATCAAGGACCTTTGGATGATATCTTCACTGTCAAGAGGA 120  
 Oy 795 TGAATTAAGTGTCAACCCAGCTGGTGGGAGAGGATGAACTACTGATTAACCTAC 854  
 Db 121 TGAATTAAGTGTCAACCCAGCTGGTGGGAGAGGATGAACTACTGATTAACCTAC 180  
 Oy 855 AGCTGGTGTGGAGTGCTGCTGTCTCAACATTTTCAAACTGTTTCCCCAGAAAGT 914  
 Db 181 AGCTGGTGTGGAGTGCTGCTGTCTCAACATTTTCAAACTGTTTCCCCAGAAAGT 240  
 Oy 915 AAGAATATTGTTATTTGCTGCTAGTGTGGAATGTAGACTTAACCTCCCTCACTATTC 974  
 Db 241 AAGAATATTGTTATTTGCTGCTAGTGTGGAATGTAGACTTAACCTCCCTCACTATTC 300  
 Oy 975 GGTGAAGCAGCTGAAGGCCAGCTTCTTATCAGTCTGTTTCTGTTTAA 1023  
 Db 301 GGTGAAGCAGCTGAAGGCCAGCTTCTTATCAGTCTGTTTCTGTTTAA 349

RESULT 13  
 ABAL5709/C  
 ID ABAL5709 standard; DNA; 20892 BP.  
 XX  
 AC ABAL5709;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Human nervous system related polynucleotide SEQ ID NO 8040.  
 XX  
 KW Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
 KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antileuc; anticonvulsant; antifungal;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200159063-A2.  
 XX  
 PD 16-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01334.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 21-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.



CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.

XX  
SQ Sequence 861 BP; 257 A; 167 C; 240 G; 191 T; 6 other;

Query Match 19.6%; Score 200.2; DB 21; Length 861;  
Best Local Similarity 98.5%; Pred. No. 4.5e-52;  
Matches: 202; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 394 GATGAGTCCAGAGAAATGTTGCAAAAGAGTTACAGAGAAACAGAGGCATCATGGTA 453  
|| |||||  
Db 657 GAGATGTCAGAGAAATGTTGCAAAAGAGTTACAGAGAAACAGAGGCATCATGGTA 716  
QY 454 CATCCCAACAGAGAGCCTGCAGTCATAGCTGGACAGAGGACAATGCCCTGGAGTCTG 513  
|||||  
Db 717 CATCCCAACAGAGAGCCTGCAGTCATAGCTGGACAGAGGACAATGCCCTGGAGTCTG 776  
QY 514 AACGAGTCTCTTGGTGGATGACCTGCTGTACCTAGCTGAGGAGGAATGCTTGCT 573  
|||||  
Db 777 ACCAGGTCCTTGGTGGATGACCTGCTGTACCTAGCTGAGGAGGAATGCTTGCT 836  
QY 574 GGAATAGCAATTACAGTTAAGGCTC 598  
|||||  
Db 837 GGAATAGCAATTACAGTTAAGGCTC 861

## RESULT 15

AAV74370  
ID AAV74370 standard; DNA; 31096 BP.

XX AC AAV74370;

XX DT 16-MAR-1999 (first entry)

XX STaphylococcus aureus contig SEQ ID #59.

DE Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.

XX Staphylococcus aureus.

PH Key Location/Qualifiers  
FT misc\_feature 1201..1260

FT /tag= a  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 3001..3060

FT /tag= b  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 4801..4860

FT /tag= c  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 6601..6660

FT /tag= d  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 8401..8460

FT /tag= e  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 10201..10260

FT /tag= f  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 12001..12060

FT /tag= g  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 13801..13860

FT /tag= h  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 15601..15660

FT /tag= i  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 17401..17460

FT /tag= j  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 19201..19260

FT /tag= k  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 21001..21060

FT /tag= l  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 22801..22860

FT /tag= m  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 24601..24660

FT /tag= n  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 26401..26460

FT /tag= o  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

FT misc\_feature 28201..28260

FT /tag= p

[illegible]

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 06:21:15 ; Search time 1731 Seconds  
(without alignments)  
867.232 Million cell updates/sec

Title: US-09-889-609B-9  
Perfect score: 1023  
Sequence: 1 atgtgtctcagttatgc.....atcagtcgtttctgtttaa 1023

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues 2085038

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	143.4	14.0	31096	7	US-08-781-986A-59
2	89.4	8.7	1830121	9	US-10-329-960-1
3	77.4	7.6	105184	9	US-09-847-513A-1
4	70.6	6.9	954	10	US-09-974-300-4749
5	69.6	6.8	930	9	US-09-738-626-1091
6	69.6	6.8	3309400	9	US-09-738-626-1091
7	57	5.6	1545	9	US-09-942-891-1
8	57	5.6	1545	9	US-09-942-891-5
9	57	5.6	1545	9	US-09-942-891-7
10	57	5.6	1545	9	US-09-942-891-8
11	49.4	4.8	1251	10	US-09-974-300-264
12	48.6	4.8	536165	9	US-09-939-964-1
13	48.4	4.7	926	10	US-09-974-300-4809
14	46.2	4.5	15249	7	US-08-781-986A-102
15	46	4.5	2698	10	US-09-939-980-207
16	38.6	3.8	1655	9	US-10-234-432-49
17	38.6	3.8	1855	9	US-10-234-432-47
18	38.6	3.8	2656	9	US-10-234-432-50
19	38.6	3.8	2675	9	US-10-234-432-46

20	38.4	3.8	1379	9	US-10-098-841-197	Sequence 197, App
c 21	37.2	3.6	1827	10	US-09-974-300-1928	Sequence 1928, Ap
c 22	37.2	3.6	1833	10	US-09-921-823-18	Sequence 18, Appl
c 23	37.2	3.6	1833	10	US-09-921-823-21	Sequence 21, Appl
c 24	36.6	3.6	1200	10	US-09-921-823-3	Sequence 3, Appl
25	35.4	3.5	1034	9	US-10-234-432-51	Sequence 51, Appl
26	35	3.4	928	10	US-09-974-300-271	Sequence 271, App
c 27	35	3.4	2000	9	US-09-938-842A-2742	Sequence 2742, Ap
c 28	35	3.4	2000	9	US-09-938-842A-3356	Sequence 3356, Ap
29	34.8	3.4	427	10	US-09-960-352-11109	Sequence 11109, A
30	34.4	3.4	915	9	US-10-169-048-59	Sequence 59, Appl
31	34.2	3.3	1359	9	US-09-938-842A-1627	Sequence 1627, Ap
32	34.2	3.3	118951	9	US-10-161-572-11	Sequence 11, Appl
33	33.8	3.3	763	9	US-10-153-668-28	Sequence 28, Appl
34	33.8	3.3	790	9	US-10-153-668-455	Sequence 455, App
35	33.8	3.3	880	10	US-09-822-849A-212	Sequence 212, App
36	33.8	3.3	1232	9	US-10-098-841-210	Sequence 210, App
37	33.8	3.3	3257	12	US-10-029-654-9	Sequence 9, Appl
38	33.8	3.3	3773	9	US-09-764-891-5804	Sequence 5804, Ap
39	33.8	3.3	3774	9	US-09-764-891-5803	Sequence 5803, Ap
40	33.6	3.3	486	9	US-09-918-995-32623	Sequence 32623, A
41	33.4	3.3	1274	9	US-09-983-802-75	Sequence 75, Appl
42	33.2	3.2	873	9	US-09-738-626-1101	Sequence 1101, Ap
43	33.2	3.2	936	9	US-10-198-846-5306	Sequence 5306, Ap
44	33.2	3.2	173808	12	US-10-003-806-10	Sequence 10, Appl
45	33.2	3.2	3309400	9	US-09-738-626-1	Sequence 1, Appl

#### ALIGNMENTS

##### RESULT 1

US-08-781-986A-59

; Sequence 59, Application US/08781986A

; Publication No. US20030054436A1

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 5255

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/781,986A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Benson, Bob

; REGISTRATION NUMBER: 30,446

; REFERENCE/DOCKET NUMBER: PB248pp

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 59:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 31096 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; US-08-781-986A-59

Query Match		14.0%	Score 143.4;	DB 7;	Length 31096;
Best Local Similarity		48.6%	Pred. No. 2.4e-34;		
Matches		456;	Conservative	0;	Mismatches 472; Indels 10; Gaps 2;
QY	25	TTTGCTGATGTTGAAGAAAGCTATATACACATTCGAGATTCATCCACCTCACACAGTG	84		
DB	1251	TTAGGAGATATCAAGAAAGCTAAAGCAAGCATTTAAACCATTTATTCGTGCAACACCTCTA	1310		
QY	85	CTAACAGCTCCATTTTGAATCAA---CTAACAGGGCGCAATCTTTTCTTCAATGTGA	141		
DB	1311	ATTAATCAATGATTTTAAGCCAAAGTATTAACATAAGGGAAATGATTTCTAAATATAGAA	1370		
QY	142	CTCTCCAGAAAACAGGATCTTTTAAGATTCGTGGTCTCTCAATGCCGTCAGAACTTG	201		
DB	1371	AATATGCAATTCACAGGATCTTTTAATTTAGAGCGCTAGCAAT-----NAAATTA	1423		
QY	202	GTTCCTGATGCTTTAGAAAGGAGCCGAAAGCTGTTGTTACTCACAGCAGTGGAAACCAT	261		
DB	1424	ATCACTTAACAGATGAACAAAAAGAAAAGGCATTTATCGCAGCATCTGCTGGGGAACCAT	1483		
QY	262	GGCCAGGCTCTACCTATGCTGCCAATTCGAAGGAATTCCTGCTTATATTGTTGGTGGCC	321		
DB	1484	GCACAAGTGTCTTTAACAGCTAAATTTATAGGCATTTGATCAACGATTTGTAATGCGCT	1543		
QY	322	CAGACAGCTCCAGACTGTAAAAAATTGCAATACAAGCCTACGAGCGTCAATTTGTATAC	381		
DB	1544	GAACAGCACCAACAGCAACCAACCAACAAAGGCTATGGGCAAGGTTATTTTA	1603		
QY	382	TGTGAACCTAGTATGAGTCCAGAGAAATGTTGCAAAAAGTTACAGAAGAAACAGAA	441		
DB	1604	AAAGGTAAAAAATTTTAACGAACTAGACTTTATATGAAGAATTTAGCGAAAGAAATGGC	1663		
QY	442	GGCATCATGTATACATCCCAACAGGAGCGCTGCGATGATAGCTGCAAGGGGAACTGATG	501		
DB	1664	ATGCAATCGTTCATCCATATGACATAGATTTGTAATGCGAGCGCCAAAGCAATTTGTT	1723		
QY	502	CTGGAAGTGTCAACAGGTTCTTTTGGTGGATGTCACCTGGTGGTACCTGTAGTGAGGA	561		
DB	1724	TTAGAAATTTAGATGATATTGGAATGTGAATACAGTCATCTACCGTTGGCGGTGGA	1783		
QY	562	GGAATGCTTCTCGAATACCAATTTACAGTTAAGCTCTGNAACCTAGTGTGAAGGTATAT	621		
DB	1784	GGATTAAATGCGAGTATGCGCCGCAATTAATCAATTTAAACCTTCAATTCATATATC	1843		
QY	622	GCTGCTGAACCCCTCAATGAGATGACTGCTACCGTCCAAAGCTGAAGGGGAACTGATG	681		
DB	1844	GGTGTTCATCTCAGATGTTTCATGGTATGGCTGAGTCTTTCTATAGAGAGATTTAACT	1903		
QY	682	CCCAATCTTTATCCTCCAGAAACCATAGCAGATGGTGTCAATPCCAGCATTTGGCTTGAAC	741		
DB	1904	GAACATCGAGTGGATAGCAATAGCAGATGGTGTGATGTAAGTTCTCTGTGAACAA	1963		
QY	742	ACCTGGCTATTATCAGGACCTTTGGATGATATCTTCACTGTCACAGGATGAAAT	801		
DB	1964	ACATATGAGTAGTTAAACATTTAGTAGATGATTTATTTCTTTGTTACTGAAGAAGAAAT	2023		
QY	802	AAGTGTGCAACCCAGCTGTTGGGAGAGGATGAAACTACTCATTCGAACCTACAGCTGT	861		
DB	2024	GAACATGCTATGAAGATTTAATGCAGCGTGCACAAATTTATCTGAAGGTGCAGCGCA	2083		
QY	862	GTTGGAGTGGCTGCTGTGTCTCAACATTTTCAAACTGTTTCCCGCAAGATTAAGAAC	921		
DB	2084	TTACCAACAGCTGCAATTTTAAGTGGAAAAATAAACAATAAATGGCTTCAAGATAAAAT	2143		
QY	922	ATTTGATTTGCTCAGTGGTGAATGTAGACTTTAAC	959		
DB	2144	GTTGTTGCATTTAGTTTTCAGGCGGGAATGTTGACTTTAAC	2181		

RESULT 2  
US-10-329-960-1  
; Sequence 1, Application US/10329960  
; Publication No. US20030099277A1

GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, F  
; FILE REFERENCE: PB186PI  
; CURRENT APPLICATION NUMBER: US/10/329,960  
; CURRENT FILING DATE: 2003-01-02  
; PRIOR APPLICATION NUMBER: US 09/643,990  
; PRIOR FILING DATE: 2000-08-23  
; PRIOR APPLICATION NUMBER: US 08/487,429  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/426,787  
; PRIOR FILING DATE: 1995-04-21  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1830121  
; TYPE: DNA  
; ORGANISM: Haemophilus influenzae  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (4747)..(4747)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (9921)..(9921)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (10150)..(10150)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (29298)..(29298)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36543)..(36543)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36551)..(36551)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36636)..(36636)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (40810)..(40810)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44416)..(44416)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44905)..(44905)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44975)..(44975)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (45593)..(45593)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (45732)..(45732)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature

```
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)

LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)

Query Match 8.7%; Score 89.4; DB 9; Length 1830121;
Best Local Similarity 46.4%; Pred. No. 8.7e-16;
Matches 291; Conservative 0; Mismatches 336; Indels 0; Gaps 0;

QY 253 GGAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGCAATTCCTGCTTATATT 312
Db 794225 GGTAACCATGGCCAGGCTGCGCATTTATCGCGCAACAATTTAGGCTTTAAAGCATTAATT 794284

QY 313 GTGGTGGCCCGACAGCTCCAGACTGTAAACCAATTTGCAATACAAAGCTACGGAGCGTCA 372
Db 794285 GTTATGCCACAAACACCCCAAGCAATTAAGTGGATGCGTGGTTTGGTGGTGAG 794344

QY 373 ATGTATACGTGAACCTAGTGTAGTCCAGAGAAATTTCCAAAAGAGTTACAGAA 432
Db 794345 GTGTTGTCACGGTGTAAATTTTCGATGAAGCCCAAGCAATTCGAGCTTCAAAA 794404

QY 433 GAACAGAGGCGATCATGTTGATACATCCCAACGAGCGCTCAGTAGCTGCACAGGG 492
Db 794405 GAAAAAACATGACATTTATTCCACCATTCGATCATCCATTTAGTAGTCGAGCAAGGC 794464
```





## GENERAL INFORMATION:

APPLICANT: Gruys, Kenneth James  
APPLICANT: Mitsky, Timothy Albert  
APPLICANT: Kishore, Ganesh Murthy  
APPLICANT: Slater, Steven Charles  
APPLICANT: Padgett, Stephen Rogers  
APPLICANT: Stark, David Martin  
TITLE OF INVENTION: Methods of Optimizing Substrate Pools and Biosynthesis of Poly-beta-hydroxybutyrate-co-poly-beta-hydroxyvalerate in Bacteria and Plasmids (WOBT:155--3)  
FILE REFERENCE: 11899.0155.DVUS02 (WOBT:155--3)  
CURRENT APPLICATION NUMBER: US/09/942,891  
CURRENT FILING DATE: 2002-02-01  
PRIOR APPLICATION NUMBER: US 09/313,123  
PRIOR FILING DATE: 1999-05-17  
PRIOR APPLICATION NUMBER: US 08/673,388  
PRIOR FILING DATE: 1996-06-28  
PRIOR APPLICATION NUMBER: US 08/628,039  
PRIOR FILING DATE: 1996-04-04  
PRIOR APPLICATION NUMBER: US 08/614,877  
PRIOR FILING DATE: 1996-03-13  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patent in version 3.1  
SEQ ID NO: 1  
LENGTH: 1545  
TYPE: DNA  
ORGANISM: Escherichia coli  
US-09-942-891-1

Query Match 5.6%; Score 57; DB 9; Length 1545;

Best Local Similarity 43.6%; Pred. No. 2.3e-07;

Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

QY 216 AGAAGGAGCGGAAAGCTGTTGTTACTCACAGAGTGGAACCATGCCAGGCTCTCAC 275  
DB 225 AGACAGAAAGCGCAGCGCTGATCAGCTCTCTCGGGTAACACAGCGCGCGGTGCG 284  
QY 276 CTATGCTGCCAAATTTGGAAGGAATTCCTCTTATATTTGTTGTCGCCCCAGACAGCTCCAGA 335  
DB 285 GTTTTCTTCTGCGCGGTAGGCGTGAAGGCCCTGATCGTTATGCCAACGCCACCGCGGA 344  
QY 336 CTGTAAAAAAGCTTGCATACAGCCTACGGAGCGTCAATTTGTTACTGTGAACCTAGTGA 395  
DB 345 CATCAAGTTCGACCGGCTCGCGGCTTCGGCGGCAAGTGTCTGCCAGCGCGCAACTT 404  
QY 396 TGAGTCCAGAGAAATTTGCAAAAAGAGTTACAGAAAGAGTTACAGAAAGAGATCATGTACA 455  
DB 405 TGATGAAGCAACCGCAAGCGATCGAATGTACAGAGAGGAGGTTTCACTGGGTGCC 464  
QY 456 TCCCAACAGGAGCCTGCAGTGATAGCTGGACAAGGACAATTTGCCCTGGAAAGTCTGAA 515  
DB 465 GCCGTTCCAGACCTCGATGTTGCCAGTTCGCGCGCGCTGCGCTGGAACCTGCTCCA 524  
QY 516 CCAGGTTCTTTGGTGGATGCTAGTGGTACCTGCTAGTGGAGGAGGAACTGCTGCTGG 575  
DB 525 GCAGGAGCGCCATCTCGACCGCGTATTTGTGCGAGTTCGCGCGCGCTGCTGGCTGTG 584  
QY 576 AATGCAATTTACAGTTAAGGCTCTGAAACCTAGTGTGAAGGTATATGCTGCTGAACCTTC 635  
DB 585 CTTGGGGTGTGATCAACAACACTGATGCCGCAATCAAGATGATCGCGGTAGAGACGGA 644  
QY 636 AATGCAATTTACAGTTAAGGCTCTGAAACCTAGTGTGAAGGAGGAACTGATGCCCAATCTTTATCC 695  
DB 645 AGACTCCGCGCTGCTGAAAGCAGCGCTGGATCGCGGTGATCGCGGTGATCGCGCGGTG 704  
QY 696 TCCGAAACCATAGCAGATGTTGTTCAATCCAGAGATTTGCTTGAACACTGCTGCTATAT 755  
DB 705 AGGGCTATTTGCTGAAGCGGTAGCGGTAAAGCGCATCGGTGACGAAACCTTCCCGTTATG 764  
QY 756 CAGGAGCTTTGGGATATCTTCACTGTCTACAGGATGAAT 800  
DB 765 CCAGAGTATCTCGACGACATCATCACCGTCTGATAGATGCGAT 809

## RESULT 8

US-09-942-891-5  
Sequence 5; Application US/09942891  
Publication No. US20030028917A1  
GENERAL INFORMATION:  
APPLICANT: Gruys, Kenneth James  
APPLICANT: Mitsky, Timothy Albert  
APPLICANT: Kishore, Ganesh Murthy  
APPLICANT: Slater, Steven Charles  
APPLICANT: Padgett, Stephen Rogers  
APPLICANT: Stark, David Martin  
TITLE OF INVENTION: Methods of Optimizing Substrate Pools and Biosynthesis of Poly-beta-hydroxybutyrate-co-poly-beta-hydroxyvalerate in Bacteria and Plasmids (WOBT:155--3)  
FILE REFERENCE: 11899.0155.DVUS02 (WOBT:155--3)  
CURRENT APPLICATION NUMBER: US/09/942,891  
CURRENT FILING DATE: 2002-02-01  
PRIOR APPLICATION NUMBER: US 09/313,123  
PRIOR FILING DATE: 1999-05-17  
PRIOR APPLICATION NUMBER: US 08/673,388  
PRIOR FILING DATE: 1996-06-28  
PRIOR APPLICATION NUMBER: US 08/628,039  
PRIOR FILING DATE: 1996-04-04  
PRIOR APPLICATION NUMBER: US 08/614,877  
PRIOR FILING DATE: 1996-03-13  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patent in version 3.1  
SEQ ID NO: 5  
LENGTH: 1545  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic  
US-09-942-891-5

Query Match 5.6%; Score 57; DB 9; Length 1545;  
Best Local Similarity 43.6%; Pred. No. 2.3e-07;  
Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

QY 216 AGAAGGAGCGGAAAGCTGTTGTTACTCACAGAGTGGAACCATGCCAGGCTCTCAC 275  
DB 225 AGACAGAAAGCGCAGCGCTGATCAGCTCTCTCGGGTAACACAGCGCGCGGTGCG 284  
QY 276 CTATGCTGCCAAATTTGGAAGGAATTCCTCTTATATTTGTTGTCGCCCCAGACAGCTCCAGA 335  
DB 285 GTTTTCTTCTGCGCGGTAGGCGTGAAGGCCCTGATCGTTATGCCAACGCCACCGCGGA 344  
QY 336 CTGTAAAAAAGCTTGCATACAGCCTACGGAGCGTCAATTTGTTACTGTGAACCTAGTGA 395  
DB 345 CATCAAGTTCGACCGGCTCGCGGCTTCGGCGGCAAGTGTCTGCCAGCGCGCAACTT 404  
QY 396 TGAGTCCAGAGAAATTTGCAAAAAGAGTTACAGAAAGAGTTACAGAAAGAGATCATGTACA 455  
DB 405 TGATGAAGCAACCGCAAGCGATCGAATGTACAGAGAGGAGGTTTCACTGGGTGCC 464  
QY 456 TCCCAACAGGAGCCTGCAGTGATAGCTGGACAAGGACAATTTGCCCTGGAAAGTCTGAA 515  
DB 465 GCCGTTCCAGACCTCGATGTTGCCAGTTCGCGCGCGCTGCGCTGGAACCTGCTCCA 524  
QY 516 CCAGGTTCTTTGGTGGATGCTAGTGGTACCTGCTAGTGGAGGAGGAACTGCTGCTGG 575  
DB 525 GCAGGAGCGCCATCTCGACCGCGTATTTGTGCGAGTTCGCGCGCGCTGCTGGCTGTG 584  
QY 576 AATGCAATTTACAGTTAAGGCTCTGAAACCTAGTGTGAAGGTATATGCTGCTGAACCTTC 635  
DB 585 CTTGGGGTGTGATCAACAACACTGATGCCGCAATCAAGATGATCGCGGTAGAGACGGA 644  
QY 636 AATGCAATTTACAGTTAAGGCTCTGAAACCTAGTGTGAAGGAGGAACTGATGCCCAATCTTTATCC 695  
DB 645 AGACTCCGCGCTGCTGAAAGCAGCGCTGGATCGCGGTGATCGCGGTGATCGCGCGGTG 704  
QY 696 TCCGAAACCATAGCAGATGTTGTTCAATCCAGAGATTTGCTTGAACACTGCTGCTATAT 755  
DB 705 AGGGCTATTTGCTGAAGCGGTAGCGGTAAAGCGCATCGGTGACGAAACCTTCCCGTTATG 764





	Qy	475	GTGATAGCTGCACAAAGGACAATAATGCCCTCGGAAGTGGTGCAACCAGATTCCCTTTGGTGGAT	534
	Dd	410455	ATCATCGCGCGCAGCACGAACCGTCGGTCTTGAGATCGTTGAGGCGCATGCCGACGTCGCG	410514
	Qy	535	GCACTGGTGCTTAGCTGTAGGTGGAGGAGAATGCTTGCTGGAATACCAATTACAGTTAAG	594
	Dd	410515	ATGTAATCTACTGCTTCCACTGTCGGGTGGCGGCTTGCGTCGAGCGGCTTGACAGCAGCGTGAAG	410574
	Qy	595	GCTCTGAAAACCTAGTGTGAAGGTATATGCTG	625
	Dd	410575	GCACTGGCGCTCATGCGAGGATCATCGGT	410605
RESULT 13				
:	US-09-974-300-4809			
:	: Sequence 4809, Application US/09974300			
:	: Patent No. US20020146721AI			
:	: GENERAL INFORMATION:			
:	: APPLICANT: Berkla, Randy M.			
:	: TITLE OF INVENTION: Methods For Monitoring Multiple Gene			
:	: FILE REFERENCE: 10085.500-US			
:	: CURRENT APPLICATION NUMBER: US/09/974,300			
:	: CURRENT FILING DATE: 2001-10-05			
:	: PRIOR APPLICATION NUMBER: 09/680,598			
:	: PRIOR FILING DATE: 2000-10-06			
:	: PRIOR APPLICATION NUMBER: 60/279,526			
:	: PRIOR FILING DATE: 2001-03-27			
:	: NUMBER OF SEQ ID NOS: 8481			
:	: SOFTWARE: FastSEQ for Windows Version 4.0			
:	: SEQ ID NO 4809			
:	: LENGTH: 926			
:	: TYPE: DNA			
:	: ORGANISM: Bacillus clausii			
:	: FEATURE:			
:	: NAME/key: misc_feature			
:	: LOCATION: (1)...(926)			
:	: OTHER INFORMATION: n = A,T,C or G			
:	US-09-974-300-4809			

Query Match	4.7%	Score 48.4	DB 10	Length 926
Best Local Similarity	44.5%	Pred. No. 9.5e-05		
Matches 383	Conservative 0	Mismatches 456	Indels 21	Gaps 4
Qy	31	GATGTTGAAAAGCTCATATCAACATTCGAGATTCTATCCACCTTCACACGAGTCTGCTRACA	90	
Db	6	GATATTATTCGAGCAAAATCAACAATTAAGATGTCGTGAGCGCATACACCTTTCGAAAA	65	
Qy	91	AGCTCCATTTTGAATCAACTAACAGGCGCAATCTTTTCTTCAAATGTGAACCTTCCAG	150	
Db	66	GATCAGGTTTGTGTGAACGATATGATGCAACAGTTTACTTTAAACGGGAAGACTTGC	125	
Qy	151	AAAACAGGATCTTTTAAGATTCGTGGTGCTCTCAATGCCGTCAACAAGTTGGTTCCTGAT	210	
Db	126	GTTGTCAGATCTTTTAAATTCGGGGCGC-----CTATTACCAATTTCTTCGTGAT	177	
Qy	211	GCTTTAGAAAGAGCGCAAGCTGTTGTTACTCACAGCAGTGGAAACCATGGCCAGCT	270	
Db	178	CAAAAGAAGAACTTGCCGCTGGC-GTTGTATGCGCAAGCGCTGAAACCATGCACAGG	236	
Qy	271	CTCACCTATGCTGCCAAATTTGGAAGGAATTCCTGCTTATATTGTGTGGTGGCCAGACAGCT	330	
Db	237	GTTGCGCTATTTCATGCGCGCGCTTTTAAAGTCAAAAGGTGTATTATTTTCATGCCGACTACC	296	
Qy	331	CACAGCTGTAAAAAACTTGCAAATACAAGCCTACGGAGCGTCAATTGTATATTCTGTGAACCT	390	
Db	297	CCAAAGCAAAAGTCGCACAAAGTGAATTTTGGCAGGGGACTATGTAGCGTCAGGTTA	356	
Qy	391	AGTGATGAGTCCAGAGAAAATGTT-----GCAAAAAAGAGTTACAGAAAGAACGAA	441	
Db	357	ATTGGCGATACGTTTGTATGATCTTATGCGGACGCCATTCAATATTGCGAAGACAGGAA	416	

Qy	442	GGCATCATGGTACATCCCAACGAGGAGCCTGCAGTGATAGCTGGACAAGGGACAATTTGCC	501
Db	417	ATGACGTTTCATCCATCCATTTTAAACCAAGACAAGATGATTTGCGGGGCAAGGAACAGTCGGA	476
Qy	502	CTGGAAGTGCTGA---ACCAGGTTCCTTTGGTGGATGCACCTGGTGGTACCTGTAGCTGGA	558
Db	477	CTTGAATTTAATGATGACATCGAGGAACGCCAGATTTTGTGTTTCATCAATCGCGCGC	536
Qy	559	GGAGGAATGCTTGCTGGGAATAGCAATTTACAGTTAAAGGCTCTCGAAACCTAGTGTGAAGGTA	618
Db	537	GCTGGGCTTATTAGTGGGATGGCCACGATATAATAAAAGCGTTTAGCCCAACTACGAAAAATG	596
Qy	619	TATGCTGCTGAACCCCTCAATATGCGACATGACTCTACCAGTCCAAAGCTGAAGGGGAAACTG	678
Db	597	ATTGGTTGTAACCTGCTGGTGCTGCATCCATGACGGAATCAATTTAAGCAAGGCAACGTT	656
Qy	679	ATGCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTCAAATCCAGCATTTGCTTTG	738
Db	657	GTTGAGTTAGACGAATCGACAAATTTGTCGATGGAGCGGCTGTGNAAAAAATCGCGGAC	716
Qy	739	AACACCTGGCCTATTATCAGGGACCTTGTGGAGATATATCTTCACCTGTACAGAGAGTGAA	798
Db	717	AAAACACTGGAAATTTGCAAAAGAACTGCTAGATGACATTTATTTGTGTGCCAAGGCAAA	776
Qy	799	ATTAAGTGTGCAACCCACCTGGTGTGGGAGAGGATCAAACTACTCATTTGAACCTACAGCT	858
Db	777	ATTTCCAGGACGATTTTAAACCTTTTACAATGAGAATGCGCTAGTTGCCGAACCGCAGG	836
Qy	859	GGTGTGGAGTGGGTGCTGT 878	
Db	837	GCTATGCCATTGCGGCTCT 856	

RESULT 14  
 US-08-781-986A-102  
 ; Sequence 102, Application US/08781986A  
 ; Publication No. US20030054436A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles Kunsch  
 ; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 5255  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/781,986A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Benson, Bob  
 ; REGISTRATION NUMBER: 30,446  
 ; REFERENCE/DOCKET NUMBER: PB248PP  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 102:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 15249 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double

```

; TOPOLOGY: linear
US-08-781-986A-102

Query Match          4.5%; Score 46.2; DB 7; Length 15249;
Best Local Similarity 45.2%; Pred. No. 0.0031;
Matches 364; Conservative 0; Mismatches 408; Indels 33; Gaps 4;

QY 31 GATGTTGAAAAGCTCATATCAACATTCGAGATTCATCCACCTCACCAGTGTCTAACCA 90
DB 13181 GATATCGATGAGGCAATTTTAAAGACTTTAAAGATATTGCAAGAAACACCTTTACAATTA 13240
QY 91 AGCTCCATTGTAATCAACTAACAGGCGCAATCTTTCTTCATCAATGTGAACCTTCCAG 150
DB 13241 GACCATTAATCTCAAAAGTATGATTGTAAGTCTTATTAACACGAGAAGATTACAA 13300
QY 151 AAAACAGGATCTTTTAAAGATTCGTGCTCTCAATGCCGTCAGAACCTTGGTTCCGTGAT 210
DB 13301 TGGGTACGTTCTTTAAATTAAGAGGTGCTTACACGCTATTTCTGTTTATCAGATGAA 13360
QY 211 GCTTTAGAAAGGAGCGCAAGCTGTTGTTTACTCACAGAGTGGAACACCATGGCCAGGCT 270
DB 13361 GCT-----AAAAGTAAAGGTATTACATGTGCAAGTCAGGTAATCATGCTCAAGGT 13411
QY 271 CTCACCTATGCTGCCAAATTTGGAAGAAATTCCTGCTTATATGTTGGTGCCCCAGACAGCT 330
DB 13412 GTTGCTTATACAGCTAAACAACTTAATTTAAACGCTGTTTATCTTTATGCGCAGCTACATA 13471
QY 331 CCAGACTGTAAAAAACTTGCATATACAGCCTACGAGCGCTCAATGTGTATCTGTGAACCT 390
DB 13472 CCITTCACAAAGGTAAATCAAGTAAAGTCTTTGGAAATAGTAACTGTTGAAGTTGTACT 13531
QY 391 AGTGATGAGTCCAGAGAAAA-----TGTTGCAAAAAGAGTTACAGAGAAGAAACAGAA 441
DB 13532 ACTGGTGATACATTTGATCACTGTTTAGCTGAAGCTTTAACTTATACAAAGTGAACATCA 13591
QY 442 GGATCATGTATACATCCCAACGAGGCGCTGAGTATGATGCTGCAAGGAGCAATGGCC 501
DB 13592 ATGAACCTTTATAGATCCATTCATTAATGTTTCATACAAATTTCTGCAAGGTAGCTTGT 13651
QY 502 CTGGAAGTGTGAACAGG-----TTCCCTTTGGTGATGCTGCTGTTGACCT 549
DB 13652 AAAGAATGCTAGACAAAGCAAGCTGCAATGTTAACTTTGATTTATCTATTGCGCGCA 13711
QY 550 GTAGGTGGAGGAGAAATGCTGTGTTGGAATAGCAATTAAGCTTAAGGCTCTGAAACCTAGT 609
DB 13712 ATTGGTGTGCGGTTTAAATTTCAAGGTATTAGTACTTACTTTAAACCTATTCACTACC 13771
QY 610 GTCAAGGTATATCTGCTGAACCTCAATGCAAGTCACTGCTACAGCTCCACGCTGAAG 669
DB 13772 ACGAAATTAATAGGTGTTGAACCTTCAGGTGCAAGTAGTATGATGAATCTGTTGGTA 13831
QY 670 GGAACACTGATGCCCAATCTTTATCTCTCCAGAAACCATAGCAGATGGTG---TCAAAATCC 726
DB 13832 AATAATCAGGTAGTACATGCTTATATCGATTAATTTGTCAGCGTGCATCTGAGCT 13891
QY 727 AGCATGGCTTGAACACCTGGCCCTATTATCAGGACCTTTGTGGATGATATCTTCACTGTC 786
DB 13892 AGATTTGGCGATATTACATTTGAAATTTGCAAAAAGAAATGCTAGATGATTACGTTCAAGTA 13951
QY 787 ACAGAGGATCAAAATTAAGTGTGCA 811
DB 13952 GATGAAGGTGCAGTTTGTGTTCTAGGA 13976

RESULT 15
US-09-939-980-207/c
; Sequence 207, Application US/09939980
; Patent No. US20020082234A1
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
;            Burdham, Martin
;            Hodgson, John
;            Knowles, David
;            Lonetto, Michael
;            Nicholas, Richard
;            Pratt, Julie
;            Reichard, Richard
;            Rosenberg, Martin
;            Ward, Judith

TITLE OF INVENTION: NO. US20020082234A1el Prokaryotic Polynucleotides,
Polypeptides and Their Uses

NUMBER OF SEQUENCES: 534
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/939,980
FILING DATE: 27-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/936,165
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50549
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 207:
SEQUENCE CHARACTERISTICS:
LENGTH: 2698 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 207:
US-09-939-980-207

Query Match          4.5%; Score 46; DB 10; Length 2698;
Best Local Similarity 46.5%; Pred. No. 0.0011;
Matches 231; Conservative 0; Mismatches 248; Indels 18; Gaps 2;

QY 31 GATGTTGAAAAGCTCATATCAACATTCGAGATTCATCCACCTCACCAGTGTCTAACCA 90
DB 575 GATATCGATGAAGCAATTTTAAAGACTTTAAAGATATTGTCAGAAAGAACACCTTTACAATTA 516
QY 91 AGCTCCATTGTAATCAACTAACAGGCGCAATCTTTCTTCAAAATGTGAACCTCTTCCAG 150
DB 515 GACCATTAATCTCAAAAGTATGATTGTAAGTTTATTTAAACAGAGAAGATTACAA 456
QY 151 AAACAGGATCTTTTAAAGATTCGTGCTCTCAATGCGCTCAGAGCTGTTGGTTCGTGAT 210
DB 455 TGGGTACGTTCTTTTAAATTAAGAGGTGCTTACACGCTATTTCTGTTTATCAGATGAA 396
QY 211 GCTTTAGAAAGGAGCGCAAGCTGTTGTTACTCACAGCAGTGAACCATGCCAGGCT 270
DB 395 GCT-----AAAAGTAAAGGTATTACATGTCGAGTGCAGGTATCATGCTCAAGGT 345
QY 271 CTCACCTATGCTGCCAAATTTGGAAGAAATTCCTGCTTATATTTGTTGGTCCCCAGACAGCT 330
DB 344 GTTGCTATACAGCTAAACAACTTAATNTAAAGCTGTTATCTTTATGCCAGCTACTACA 285
QY 331 CCAGACTGTAAAAAACTTGCATATACAGCTCAGCA-----CGGTCAATTGTATAC 381
DB 284 CCTTNACAAAGGCTAAATCAAGTAAAGTTCTTTGGAATAGTAACGTTGAAGTTGTACTC 225
```

QY	382	TGTGAACCTAGTGATCGAGTCCAGAGAAAATGTTGCAAAAAGAGTTTACAGAAGAAACAGAA	441
Db	224	ACTGGTGATACATTTCGATCNCCTGTTTAGCTGAAGCTTTAACTTTATACAAGTGAACATCAA	165
QY	442	GGCATCATGGTACATCCCAACCCAGGAGCCTGCAGTGATAGCTGGACAAGGGACNATTGCC	501
Db	164	ATGAACCTTTATAGATCCCATTCATAATGTTTCATACAATTTCTGGACAAGGTACGCTTGCT	105
QY	502	CTGGAAGTGCTGAACCA	518
Db	104	AAAGAAATGCTAGACA	88

Search completed: June 24, 2003, 09:40:02  
Job time : 1742 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:35:44 ; Search time 46.0677 Seconds  
(without alignments)  
983.448 Million cell updates/sec

Title: US-09-889-609B-10  
Perfect score: 1735  
Sequence: 1 MCAQVCISFADVEKAHINR.....SSITWVQKQERPAQSVSV 340

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1735	100.0	340	21 AAB07734	Amino acid sequenc
2	1735	100.0	340	22 AAU09124	Human Serine Racem
3	1735	100.0	340	22 AAE08342	Human pyridoxal-ph
4	1735	100.0	340	22 AAB94477	Human protein sequ
5	1731	99.8	340	22 AAM50262	Human serine racem
6	1727	99.5	340	22 AAG78808	Serine/threonine d
7	1582.5	91.2	339	21 AAB07731	Amino acid sequenc
8	1062.5	61.2	228	23 ABB89713	Human polypeptide
9	450	25.9	469	22 ABB63708	Drosophila melanog
10	414.5	23.9	1181	22 ABB24298	Novel human diagno

11	368	21.2	502	20 AAY32941	Mutant threonine d
12	368	21.2	502	20 AAY05705	Feedback insensiti
13	368	21.2	532	20 AAY32943	Mutant threonine d
14	368	21.2	532	20 AAY05707	Feedback insensiti
15	368	21.2	539	20 AAY32942	Mutant threonine d
16	368	21.2	539	20 AAY05706	Feedback insensiti
17	368	21.2	545	20 AAY32947	Mutant threonine d
18	368	21.2	545	20 AAY05711	Feedback insensiti
19	368	21.2	592	20 AAY32939	Mutant threonine d
20	368	21.2	592	20 AAY32951	Wild type threonin
21	368	21.2	592	20 AAY05702	Arabidopsis wild-t
22	368	21.2	592	20 AAY05703	Feedback insensiti
23	368	21.2	600	20 AAY32952	Mutant threonine d
24	368	21.2	609	20 AAY32940	Mutant threonine d
25	368	21.2	609	20 AAY05704	Feedback insensiti
26	367	21.2	424	22 AAG81976	S. epidermidis ope
27	367	21.2	424	23 ABP39031	Staphylococcus epi
28	363	20.9	590	20 AAY32950	Mutant threonine d
29	363	20.9	592	20 AAY32948	Mutant threonine d
30	361.5	20.8	422	23 ABB48174	Listeria monocytog
31	351.5	20.3	416	23 ABB54559	Lactococcus lactis
32	349.5	20.1	310	22 AAG90837	C glutamicum prote
33	349.5	20.1	310	22 AAB79703	Corynebacterium gl
34	348.5	20.1	441	15 AAR54223	L.lactis branched
35	312.5	18.0	436	16 AAR94690	Threonine dehydrat
36	311.5	18.0	436	22 AAG92074	C glutamicum prote
37	311	17.9	621	22 ABG30299	Novel human diagno
38	310.5	17.9	436	16 AAR94686	Threonine dehydrat
39	309.5	17.8	423	22 AAB79762	Corynebacterium gl
40	309.5	17.8	436	16 AAR94687	Threonine dehydrat
41	309.5	17.8	436	16 AAR94688	Threonine dehydrat
42	307.5	17.7	436	16 AAR94689	Threonine dehydrat
43	307.5	17.7	436	16 AAR94691	Threonine dehydrat
44	305	17.6	316	22 ABB66014	Drosophila melanog
45	279.5	16.1	340	22 AAU23238	Novel human enzyme

ALIGNMENTS

RESULT 1  
AAB07734  
ID AAB07734 standard; Protein; 340 AA.  
XX  
AC AAB07734;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE Amino acid sequence of a human serine racemase polypeptide.  
XX  
KW Serine racemase; N-methyl-D-aspartate receptor; neural death;  
KW neural dysfunction; NMDA receptor; Parkinson's disease;  
KW Huntington's disease; motor neurone disease; Alzheimer's disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200043526-A1.  
XX  
PD 27-JUL-2000.  
XX  
PF 18-JAN-2000; 2000WO-US00938.  
XX  
PR 19-JAN-1999; 99US-0116333.  
PR 21-JUL-1999; 99US-0144839.  
XX 28-JUL-1999; 99US-0145953.  
XX  
XX (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
XX Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;  
PI Ferris CD;  
XX WPI; 2000-482915/42.  
DR N-PSDB; AAA59299.



XX AAE08342;  
XX 15-NOV-2001 (first entry)  
XX Human pyridoxal-phosphate dependent enzyme 22406 protein.  
XX  
KW Human; pyridoxal phosphate dependent enzyme; nootropic; neuroprotective;  
KW anticonvulsant; cerbroprotective; cardiant; vasotropic; gene therapy;  
KW epilepsy; Alzheimer's disease; chromosome 17; serine racemase; stroke;  
KW behavioural change; neurodegenerative disorder; schizophrenia; atresia;  
KW rheumatic heart failure; circulatory disorder; hepatic injury; jaundice;  
KW lung disorder; nodular hyperplasia; stenosis; skeletal muscle disorder;  
KW tumour; rhabdomyosarcoma; dermal fibroblast disorder.  
XX  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Modified-site 8..11  
FT Domain /label= Casein\_kinase\_II\_phosphorylation\_site  
FT 19..315  
FT /note= "Pyridoxal-phosphate dependent enzyme family  
FT domain"  
FT Modified-site 38..40  
FT /label= Protein\_kinase\_C\_phosphorylation\_site  
FT Modified-site 47..60  
FT /note= "Serine/threonine dehydratases pyridoxal-phosphate  
FT attachment site"  
FT Modified-site 54..56  
FT /label= Protein\_kinase\_C\_phosphorylation\_site  
FT Modified-site 59..64  
FT /label= N\_myristoylation\_site  
FT Modified-site 88..93  
FT /label= N\_myristoylation\_site  
FT Modified-site 109..112  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Modified-site 140..143  
FT /note= "cAMP and cGMP-dependent protein kinase  
FT phosphorylation site"  
FT Domain 176..197  
FT /label= Transmembrane\_domain  
FT Modified-site 187..192  
FT /label= N\_myristoylation\_site  
FT Modified-site 196..198  
FT /label= Protein\_kinase\_C\_phosphorylation\_site  
FT Modified-site 203..205  
FT /label= Protein\_kinase\_C\_phosphorylation\_site  
FT Modified-site 212..215  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Modified-site 235..238  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Modified-site 239..244  
FT /label= N\_myristoylation\_site  
FT Modified-site 261..264  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Modified-site 287..292  
FT /label= N\_myristoylation\_site  
FT Domain 308..326  
FT /label= Transmembrane\_domain  
XX  
XX WO200160987-A1.  
XX  
XX 23-AUG-2001.  
XX  
XX 20-FEB-2001; 2001WO-US053365.  
XX  
XX 17-FEB-2000; 2000US-0183208.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Meyers RA, Rudolph-Owen LA;  
XX  
XX WPI; 2001-529909/58.  
XX

DR N-PSDB: AAD14461.  
XX Novel polypeptide of the human pyridoxal phosphate dependent family  
PT useful in screening and detection assays and for treatment, e.g. of  
PT epilepsy and Alzheimer's  
XX  
XX Claim 8; Fig 1; 12lpp: English.  
XX  
XX The present sequence is human pyridoxal phosphate dependent enzyme  
CC 22406 which is a serine racemase. Human 22406 gene is located on  
CC chromosome 17 between D17S849 and D17S796. The protein 22406  
CC is a modulator of D-serine. D serine has been shown to modify  
CC behavioural changes associated with learning, memory and convulsions.  
CC Human 22046 and compounds that modulate the expression or activity are  
CC used to treat or diagnose neurodegenerative disorders including  
CC Alzheimer's disease, schizophrenia as well as quell anxiety and  
CC epilepsy and prevent damage from stroke as well as cardiac (heart  
CC failure, rheumatic heart failure) and circulatory disorders, liver  
CC disorders (hepatic injury, jaundice), lung disorders, prostate  
CC disorders (benign enlargement, nodular hyperplasia), colon disorders  
CC (atresia, stenosis), skeletal muscle disorders (tumours-rhabdomyosarcoma)  
CC and dermal fibroblast disorders. Human 22406 cDNA is also useful in gene  
CC therapy.  
XX  
XX Sequence 340 AA;  
SQ  
Query.Match 100.0%; Score 1735; DB 22; Length 340;  
Best Local Similarity 100.0%; Pred. No. 1.6e-167;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCAQYCIISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60  
DB 1 MCAQYCIISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60  
QY 61 LNAVRSFLVPDALERPKAVVTHSSNGHGOALTYAAKLEGIPAYIVVPQTAPDCKKLAIOA 120  
DB 61 LNAVRSFLVPDALERPKAVVTHSSNGHGOALTYAAKLEGIPAYIVVPQTAPDCKKLAIOA 120  
QY 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNQPAPVIAQGGTIALEVLNQVPLVDAL 180  
DB 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNQPAPVIAQGGTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGMLAGIAITVRALKPSVKVYAAEPSNADDCYOSKLGKMLPNLYPETIADGV 240  
DB 181 VVPVGGGMLAGIAITVRALKPSVKVYAAEPSNADDCYOSKLGKMLPNLYPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLVDVDFITVTEDEIKCATOLVWERMKLLIETPTAGVGVAAVLSQHFT 300  
DB 241 KSSIGLNTWPIIRDLVDVDFITVTEDEIKCATOLVWERMKLLIETPTAGVGVAAVLSQHFT 300  
QY 301 VSPEVKNICIVLSGGNVDLTSSITWVQKAEKRPASQSVSV 340  
DB 301 VSPEVKNICIVLSGGNVDLTSSITWVQKAEKRPASQSVSV 340  
RESULT 4  
AAB94477  
ID AAB94477 standard; Protein; 340 AA.  
XX  
XX AAB94477;  
XX  
XX 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:15149.  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
XX Homo sapiens.  
XX  
XX EP1074617-A2.  
XX  
XX PD 07-FEB-2001.  
XX

PF 28-JUL-2000; 2000EP-0116126.  
 XX 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX (HELI-) HELIX RES INST.  
 XX  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 XX WPI: 2001-318749/34.  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 XX Claim 8; SEQ ID 15149; 2537pp + CD ROM; English.  
 PS  
 XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to a  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 XX Sequence 340 AA;  
 SQ  
 Query Match 100.0%; Score 1735; DB 22; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-167;  
 Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MCAOYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60  
 Db 1 MCAOYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60  
 QY 61 LNAVRLVPDALERPKAVVTHSSNGHQAITYAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
 Db 61 LNAVRLVPDALERPKAVVTHSSNGHQAITYAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
 QY 121 YGASIVCEPDSERENAKRVTEETEGIMVHPNQPENAVIAGQGTIALEVLNQVPLDAL 180  
 Db 121 YGASIVCEPDSERENAKRVTEETEGIMVHPNQPENAVIAGQGTIALEVLNQVPLDAL 180  
 QY 181 VVPVGGGMLAGIATTVKALPSVKVYAAEPSNADDCYOSKLGKMLPNLYPPETIADGV 240  
 Db 181 VVPVGGGMLAGIATTVKALPSVKVYAAEPSNADDCYOSKLGKMLPNLYPPETIADGV 240  
 QY 241 KSSIGLNTWPIIRLDVDDIFTVTEDEIKCATOLVWERMKLLIEPTAGVGVAAVLVLSQHFT 300  
 Db 241 KSSIGLNTWPIIRLDVDDIFTVTEDEIKCATOLVWERMKLLIEPTAGVGVAAVLVLSQHFT 300  
 QY 301 VSPVKNICIVLSGGNVDLTSSITVWKOERPASYQSVSV 340  
 Db 301 VSPVKNICIVLSGGNVDLTSSITVWKOERPASYQSVSV 340

Db 301 VSPVKNICIVLSGGNVDLTSSITVWKOERPASYQSVSV 340  
 RESULT 5  
 AAM50262  
 ID AAM50262 standard; Protein: 340 AA.  
 AC AAM50262;  
 XX 21-JAN-2002 (first entry)  
 DT Human serine racemase.  
 DE  
 XX Serine racemase; human; D-serine; regulation;  
 KW glutamate N-methyl-D-aspartate receptor; neurodegenerative disease;  
 KW stroke; Alzheimer's disease; Parkinson's disease; pain; myopathy;  
 KW neurotropic; neuroprotective; cerebroprotective; antiparkinsonian;  
 KW analgesic; diagnosis; therapy; screening.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT Modified-site 41..59  
 FT /note= "prosite serine/threonine dehydratase  
 FT pyridoxal-phosphate attachment site"  
 FT Modified-site 71..128  
 FT /note= "serine/threonine dehydratase  
 FT pyridoxal-phosphate attachment site"  
 FT Modified-site 154..169  
 FT /note= "cysteine synthase/cystathione beta-synthase  
 FT attachment site"  
 FT  
 PN WO2001173077-A2.  
 PD 04-OCT-2001.  
 XX 30-MAR-2001; 2001WO-EP03668.  
 XX 31-MAR-2000; 2000US-193748P.  
 PR 03-APR-2000; 2000US-194249P.  
 XX (FARB ) BAYER AG.  
 XX Ramakrishnan S;  
 XX WPI: 2001-648444/74.  
 DR N-PSDB; AAI70575.  
 XX  
 PT Polynucleotide encoding serine racemase enzyme and the enzyme useful  
 PT for screening reagents regulating the activity of the enzyme in a  
 PT neuron disease caused by over- or under-activation of glutamate  
 PT N-methyl-D-aspartate  
 XX Claim 1; Fig 2; 66pp; English.  
 PS  
 XX The present sequence is that of human serine racemase, an enzyme  
 CC that catalyses the conversion of L-serine to D-serine. Neuron  
 CC damage following various nervous system diseases is often caused by  
 CC activation of glutamate N-methyl-D-aspartate (NMDA) receptors in  
 CC the brain. This receptor is activated by the binding of D-serine.  
 CC Regulation of D-serine levels through regulation of serine racemase  
 CC may therefore prevent or minimise neuron damage in neurogenic and  
 CC myopathic disorders, neurodegenerative disorders such as Alzheimer's  
 CC disease and Parkinson's disease, and disorders leading to peripheral  
 CC and chronic pain. The invention provides serine racemase  
 CC polypeptides and methods for their recombinant production. These  
 CC polypeptides are used in claimed methods of screening for agents  
 CC that modulate or decrease serine racemase activity. A claimed  
 CC pharmaceutical composition comprising either an expression vector  
 CC that contains a serine racemase polynucleotide, or a reagent that  
 CC modulates serine racemase enzyme activity, is used to modulate  
 CC serine racemase activity in a disease, particularly neuron damage or  
 CC a neurodegenerative disease caused by the over- or under-activation

CC of the glutamate NMDA receptor.

XX Sequence 340 AA;  
SQ  
Query Match 99.8%; Score 1731; DB 22; Length 340;  
Best Local Similarity 99.7%; Pred. No. 4.1e-167;  
Matches 339; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCAQYCTSPADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLFKCELFOKTGSFKIRGA 60  
DB 1 MCAQYCTSPADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLFKCELFOKTGSFKIRGA 60  
QY 61 LNAVRSILVPDALERKPKAVVTHSSGNHGQALTAAKLEGIPAYIVVPQTAPDCCKKLAIOA 120  
DB 61 LNAVRSILVPDALERKPKAVVTHSSGNHGQALTAAKLEGIPAYIVVPQTAPDCCKKLAIOA 120  
QY 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNQEPAVIAGOGTIALEVLNQVPLVDAL 180  
DB 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNQEPAVIAGOGTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYQSKLKGKLPNLYPPETIADGV 240  
DB 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYQSKLKGKLPNLYPPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLVDDIFTVTDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFQT 300  
DB 241 KSSIGLNTWPIIRDLVDDIFTVTDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFQT 300  
QY 301 VSPEVKNICIVLGGNVDLTSSITWVKQAEPPASYSQSVS 340  
DB 301 VSPEVKNICIVLGGNVDLTSSITWVKQAEPPASYSQSVS 340

RESULT 6  
AAG78808  
ID AAG78808 standard; Protein; 340 AA.

XX AC AAG78808;  
XX DT 27-NOV-2001 (first entry)  
XX DE Serine/threonine dehydrase 37.  
XX KW Serine/threonine dehydrase 37; cancer; HIV infection; anti-HIV;  
XX cytosstatic.

OS Unidentified.  
XX CN1300824-A.  
XX PD 27-JUN-2001.

XX PF 21-DEC-1999; 99CN-0125662.

XX PR 21-DEC-1999; 99CN-0125662.

XX PA (YFU-) UNIV FUDAN.

XX PI Mao Y, Xie Y;

XX WPI: 2001-530468/59.

XX N-PSDB: AAI65020.

XX Polypeptide-serine/threonine dehydrase 37 and polynucleotide for coding  
PT this polypeptide -

XX Claim 1; Page 27 (Disclosure); 34pp; Chinese.

XX The present sequence is the protein sequence for serine/threonine  
CC dehydrase 37. The dehydrase and its coding sequence are useful for  
CC treating diseases e.g. cancer and HIV infection.

XX Sequence 340 AA;  
SQ

Query Match 99.5%; Score 1727; DB 22; Length 340;  
Best Local Similarity 99.7%; Pred. No. 1e-166;  
Matches 339; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MCAQYCTSPADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLFKCELFOKTGSFKIRGA 60  
DB 1 MCAQYCTSPADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLFKCELFOKTGSFKIRGA 60  
QY 61 LNAVRSILVPDALERKPKAVVTHSSGNHGQALTAAKLEGIPAYIVVPQTAPDCCKKLAIOA 120  
DB 61 LNAVRSILVPDALERKPKAVVTHSSGNHGQALTAAKLEGIPAYIVVPQTAPDCCKKLAIOA 120  
QY 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNQEPAVIAGOGTIALEVLNQVPLVDAL 180  
DB 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNQEPAVIAGOGTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYQSKLKGKLPNLYPPETIADGV 240  
DB 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYQSKLKGKLPNLYPPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLVDDIFTVTDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFQT 300  
DB 241 KSSIGLNTWPIIRDLVDDIFTVTDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFQT 300  
QY 301 VSPEVKNICIVLGGNVDLTSSITWVKQAEPPASYSQSVS 340  
DB 301 VSPEVKNICIVLGGNVDLTSSITWVKQAEPPASYSQSVS 340

RESULT 7  
AAB07731  
ID AAB07731 standard; Protein; 339 AA.

XX AC AAB07731;

XX DT 07-NOV-2000 (first entry)

XX DE Amino acid sequence of a murine serine racemase polypeptide.

XX KW Serine racemase; N-methyl-D-aspartate receptor; neural death;

XX KW neural dysfunction; NMDA receptor; Parkinson's disease;

XX KW Huntington's disease; motor neurone disease; Alzheimer's disease.

XX OS Mus musculus.

XX Key Location/Qualifiers

XX FT Misc-difference 337

XX FT /note= "Ser encoded by T in AAA59294"

XX WO200043526-A1.

XX PD 27-JUL-2000.

XX PF 18-JAN-2000; 2000WO-US00938.

XX PR 19-JAN-1999; 99US-0116333.

XX PR 21-JUL-1999; 99US-0144839.

XX PR 28-JUL-1999; 99US-0145953.

XX (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

XX PI Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;  
XX Ferris CD;

XX WPI: 2000-482915/42.

XX N-PSDB: AAA59294, AAA59300.

XX Mammalian serine racemase preparations, used to identify modulators  
PT which can be used to treat diseases associated with  
PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease

XX

PS Claim 7; Page 48-49; 54pp; English.

xx The present sequence represents a mammalian serine racemase, which has  
 CC a specific activity of at least 0.003 micromole L-serine/mg/hour.  
 CC The enzyme catalyses the direct racemisation of L-serine to D-serine.  
 CC D-serine appears to be an endogenous ligand of N-methyl-D-aspartate  
 CC (NMDA) receptors. The mammalian serine racemases can be used to identify  
 CC modulators, which can be used in the treatment of acute or chronic  
 CC neural death or dysfunction mediated by overactivation of N-methyl-D-  
 CC aspartate (NMDA) receptors. Overactivation of the receptors is  
 CC associated with Parkinson's disease, Huntington's disease, motor neurone  
 CC disease and Alzheimer's disease.

xx Sequence 339 AA;

Query Match 91.2%; Score 1582.5; DB 21; Length 339;  
 Best Local Similarity 89.7%; Pred. No. 5.1e-152;  
 Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;  
 QY 1 MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60  
 DB 1 MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60  
 QY 61 LNAVSLVPDALEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
 DB 61 LNAVSLVPDALEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
 QY 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNOEPAVIAGOGTIALEVLNOVPLVDAL 180  
 DB 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNOEPAVIAGOGTIALEVLNOVPLVDAL 180  
 QY 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYQSKLKGKLMPLNPETIADGV 240  
 DB 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYQSKLKGKLMPLNPETIADGV 240  
 QY 241 KSSIGLNTWPIIRDLVDVDTFTVEDEIKCATQLWGRMKLLIETAGVGVAAVLSQHFQT 300  
 DB 241 KSSIGLNTWPIIRDLVDVDTFTVEDEIKCATQLWGRMKLLIETAGVGVAAVLSQHFQT 300  
 QY 301 VSPVKKNCIVLSGGNVLDLTSSITWVKQAEKPSYQSVV 340  
 DB 301 VSPVKKNCIVLSGGNVLDLTSSITWVKQAEKPSYQSVV 340

RESULT 8  
 ABB89713  
 ID ABB89713 standard; Protein; 228 AA.

AC ABB89713;

DT 24-MAY-2002 (first entry)

DE Human polypeptide SEQ ID NO 2089.

xx Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 xx antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 xx vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 xx cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 xx neurological disease; infection; human; secreted protein.

OS Homo sapiens.

xx WO200190304-A2.

xx 29-NOV-2001.

xx 18-MAY-2001; 2001WO-US16450.

xx 19-MAY-2000; 2000US-205515P.

xx (HUMA-) HUMAN GENOME SCI INC.

xx Birse CE, Rosen CA;

xx WPI: 2002-122018/16.  
 DR N-PSDB; ABL90122.

xx Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive,  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders.

xx Claim 11; SEQ ID NO 2089; 2081pp + Sequence Listing; English.

xx The invention relates to novel genes (ABL9449-ABL90853) and proteins  
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and antagonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.

xx Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at [fp.wipo.int/pub/published\\_pct\\_sequences](http://fp.wipo.int/pub/published_pct_sequences).

xx Sequence 228 AA;

Query Match 61.2%; Score 1062.5; DB 23; Length 228;  
 Best Local Similarity 93.4%; Pred. No. 2e-99;  
 Matches 213; Conservative 2; Mismatches 12; Indels 1; Gaps 1;  
 QY 1 MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60  
 DB 1 MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60  
 QY 61 LNAVSLVPDALEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
 DB 61 LNAVSLVPDALEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
 QY 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNOEPAVIAGOGTIALEVLNOVPLVDAL 180  
 DB 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNOEPAVIAGOGTIALEVLNOVPLVDAL 180  
 QY 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYQSKLKGKLMPLNPETIADGV 228  
 DB 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYQSKLKGKLMPLNPETIADGV 227

RESULT 9

ABB63708

ID ABB63708 standard; Protein; 469 AA.

xx ABB63708;

xx 26-MAR-2002 (first entry)

xx Drosophila melanogaster polypeptide SEQ ID NO 17916.

xx Drosophila; developmental biology; cell signalling; insecticide;  
 xx pharmaceutical.

xx Drosophila melanogaster.

xx WO200171042-A2.

xx 27-SEP-2001.

xx 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
PA (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
DR N-PSDB; ABL07811.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX Disclosure; SEQ ID NO 17916; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 469 AA;  
SQ  
Query Match 25.9%; Score 450; DB 22; Length 469;  
Best Local Similarity 36.6%; Pred. No. 1.2e-36;  
Matches 117; Conservative 52; Mismatches 133; Indels 18; Gaps 8;  
QY 7 ISFADVEKAHNIIRDSIHLPVTSSILNQLTGRNLFKCELFQKTGSKIRGALNAVR 66  
DB 59 ISFHDVTSAAFLIRGGVETP-CPKSTSDLYGMELYLKKDLQYTGSPKRGARYALLS 117  
QY 67 LVPDALERPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPTAPDCKKLAIOAGV 126  
DB 118 LFE---EQRTGVISASLNHAQALCYHGWKNIPVTVVWPKAAPIMKIKCRNKYARVI 174  
QY 127 YCEPDSRENVAKRVTETEGIMVHPNOEPVAVIAGQGTIALEVLNQVPLDALVVPVG 186  
DB 175 VDGNDGGEAKSLAMRMSREGLLYNGYDHPHIMAGQGTIGLEILEQVPEPDAVVPVG 234  
QY 187 GGLAGIATVVKALPSVKVYAAEPSNADDCYQSKLKGKLMNP---LYPP---ETIADGVK 241  
DB 235 GGLIAGIATVVKALPSVKVYAAEPSNADDCYQSKLKGKLMNP---LYPP---ETIADGVK 288  
QY 242 -SSIGLNTWPIIRDLVDLFTVTEDEIKCATOLVWERMKLLLEPTAGVGVAAVLSO 300  
DB 289 VPKVGNATATAMPLIDRMVWVKEEWIAVAILRLVEEKCVCVVEGAGGAGLAAILAGH 348  
QY 301 VSPVKNICIVLGGGNDLT 320  
DB 349 L--KGRKVVVLLCGGNIDT 366  
RESULT 10  
ABG24298  
ID ABG24298 standard; Protein; 1181 AA.  
XX  
XX AC ABG24298;  
XX  
XX 18-FEB-2002 (first entry)  
DT  
XX Novel human diagnostic protein #24289.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX

OS Homo sapiens.  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
PF  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS88485.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 20; SEQ ID No 54657; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1181 AA;  
SQ  
Query Match 23.9%; Score 414.5; DB 22; Length 1181;  
Best Local Similarity 36.6%; Pred. No. 2e-32;  
Matches 102; Conservative 54; Mismatches 108; Indels 15; Gaps 7;  
QY 42 LFFKCELFQKTGSKIRGALNAVRSLVDPALERKPRVTVHSSNGHQALTYAAKLEGIP 101  
DB 15 IFLKFENMORTGSKIRGAFNKLSSLT-DA--EKRGVAVACSNAGQVSLSCAMLGID 71  
QY 102 AVTVVPTAPDCKKLAIOAGVAGSIVYCEPSDESRENVAK--RVTEETEGIMVHPNOEP 159  
DB 72 GKVVMPKGA PKSKVAATCDYSAEVL--HCDNFNDIAKVEISEIVEARIFIPYDDPKV 129  
QY 160 IAGQGTIALEVLNQVPLDALVVPVG GGLAGIATVVKALPSVKVYAAEPSNADDCYQ 219  
DB 130 IAGQGTIGLEIMEDLYDVNDVIVPIGGGGLIAGIAVAKSINPTIRVQGSENVHGVMAA 189  
QY 220 SKLKGKLMNPVLPETIADGVK--SSIGLNTWPIIRDLVDLFTVTEDEIKCATOLV 278  
DB 190 SFHSGEITTH-RTTGTADGCDVSRPGLNTYETVRELVDLVLVSEDEIRNSMIALIQ 248  
QY 279 KLLIETAGVGVAAVLSQH-----FQTVSPVKNICIV 311  
DB 249 KVVTEGAGALACALLSEAAARGYEIFEKREECRKKVLY 287

```

RESULT 11
AAV32941
ID AAY32941 standard; Protein; 502 AA.
XX
AC AAY32941;
XX
XX
DT 09-NOV-1999 (first entry)
XX
DE Mutant threonine dehydratase/deaminase protein sequence.
XX
XX Threonine dehydratase/deaminase; TD; feedback insensitive mutant;
KW molecular marker; isoleucine toxic structural analog resistance;
KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;
KW polyhydroxybutyrate; antibiotic resistance marker; mutin.
XX
XX Arabidopsis thaliana.
OS Synthetic.
XX
PN WO9941395-A1.
XX
PD 19-AUG-1999.
XX
PF 08-JAN-1999; 99WO-0500560.
XX
PR 10-JUL-1998; 98WO-US14362.
PR 17-FEB-1998; 98US-0074875.
XX
PA (DOWC ) DOW AGROSCIENCES LLC.
PA (PURD ) PURDUE RES FOUND.
XX
XX Larrinua IM, Merio DJ, Mourad GS, Paredy DR;
XX
DR WPI; 1999-527375/44.
DR N-PSDB; AAZ11199.
XX
XX New nucleic acid encoding threonine dehydratase/deaminase resistant
PT to feedback inhibition, useful as selection marker for cell
PT transformation and to impart herbicide resistance
XX
PS Claim 13; Page 102-104; 194pp; English.
XX
XX This sequence represents a mutant Arabidopsis thaliana threonine
CC dehydratase/deaminase (TD) protein of the invention. The protein is a
CC feedback insensitive mutant. The TD DNA sequence is used as molecular
CC marker (imparting resistance to toxic structural analogues of isoleucine)
CC for selecting transformed cells and to produce transformants with
CC increased levels of isoleucine (and thus better nutritional value) or of
CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for
CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also
CC TD-expressing plants permit use of the isoleucine structural analogues as
CC herbicides. The DNA sequences are alternatives for antibiotic resistance
CC markers (which are potentially harmful to the environment). Since no
CC human analog of TD exists (humans can not synthesize isoleucine), it
CC should be safe to use.
XX
SQ Sequence 502 AA;
Query Match 21.2%; Score 368; DB 20; Length 502;
Best Local Similarity 30.2%; Pred. No. 2.8e-28;
Matches 102; Conservative 65; Mismatches 129; Indels 42; Gaps 10;
QY 26 TPVLTSSILNLTGRNLFKCELFQKFGKIRGALNAVRSVLPDALKRKPAAVTHSSG 85
DB 21 SPQLAKKUSKRGVRYUKREDLQVFSFKURGYNNMVKLPADQL---AGVICSSAG 77
QY 86 NHQQUALTYAAKLEGIPAYIVVPQAPDCKKLAIQAYGASIVCEPSDESRENKRVTEE 145
DB 78 NHAQGVALSASKLGCTAVIVMPVTTPEIKQAVENIGATVLFGDSYDQAOAHK-IRAE 136
QY 146 TEGI-WVHNQEPAVTAGOGTTALEVLNQV--PLVDALVVPVGGGGMLAGIAITVYKALP 202
DB 137 EELGTPIPPFDHPDVIAGOGTVGMEITROAKGPL-HAIFVPVGGGGGLIAGIAAYKRVSP 195

```

---

```

QY 203 SVKYAAEPSNADDCYQSKLGKLMFNLYPPETI-----ADGVK-SSIGLNTWPIIR 253
DB 196 EVKIIGVEPADAN-----AMALSLLHGGERVILDQVGGFADGVAVVEGEETFRISR 246
QY 254 DLVDDIFTVTEDEIKCATQOLVWERMKLLIETAGVGAVALSQHFQTVSPVKNICIVLS 313
DB 247 NLMGDGVVLVTRDAICASIKDMFEKRNILEPAGALALAGA-BAYCKYYGLKDVNVVAITS 305
QY 314 GGNVDL-----TSSITWVKQAEPPASYQ 336
DB 306 GAMNMFDKLRIVTELAVNGRQEQEAVLATLMPKPGSFK 343
RESULT 12
AAY05705
ID AAY05705 standard; Protein; 502 AA.
XX
AC AAY05705;
XX
XX 19-JUL-1999 (first entry)
XX
DE Feedback insensitive mature threonine dehydratase/deaminase.
XX
KW Threonine dehydratase/deaminase; omr1 gene; feedback inhibition;
KW transgenic plant; selectable marker; isoleucine; mutant.
XX
OS Arabidopsis thaliana.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Misc-difference 409 /note= "Arg in wild-type enzyme"
FT Misc-difference 454 /note= "Arg in wild-type enzyme"
FT Region 396..414 /note= "regulatory region R4"
FT Region 446..464 /note= "regulatory region R6"
XX
XX WO9902656-A1.
XX
XX 21-JAN-1999.
XX
XX 10-JUL-1998; 98WO-US14362.
XX
XX 17-FEB-1998; 98US-0074875.
XX 10-JUL-1997; 97US-0052096.
XX
XX (PURD ) PURDUE RES FOUND.
XX
XX Mourad GS;
XX
XX WPI; 1999-120860/10.
XX N-PSDB; AAX25334.
XX
XX New sequences encode mutant threonine dehydratase/deaminase - which
PT is insensitive to feedback inhibition, useful as a selective marker
PT to produce transformed cells resistant to toxic isoleucine analogues
XX
XX Disclosure; Page 60-62; 120pp; English.
XX
XX This sequence represents an Arabidopsis thaliana mutant mature
CC threonine dehydratase/deaminase (TD) which, unlike the wild-type
CC enzyme (see AAY05702), is insensitive to feedback inhibition by
CC isoleucine. It lacks the chloroplast transit peptide of the
CC mutant TD precursor, and is encoded by a polynucleotide (see
CC AAX25334) that is one of 9 claimed polynucleotides (see AAX25332-40),
CC originally isolated and cloned from A. thaliana mutant line GM11b
CC (omr1/omr1), which encode a feedback insensitive TD. These
CC polynucleotides can be used to transform a wide variety of plants,
CC fungi, bacteria and yeast. The polynucleotides differ from the
CC wild-type only by 2 point mutations, which result in R499C and

```

CC R554H amino acid substitutions (numbering according to wild-type  
 CC TD) in the R4 and R6 regulatory regions. These forms of TD are not.  
 CC only insensitive to feedback inhibition by isoleucine, but are also  
 CC insensitive to structural analogues of isoleucine that are toxic to  
 CC plants and microorganisms which synthesise only wild-type TD.  
 CC Nucleotide sequences encoding mutated forms of TD can therefore be  
 CC used to create cells that are insensitive to compounds normally  
 CC toxic to cells expressing only wild-type TD enzymes, and thus may  
 CC be used to provide a biochemical selectable marker. Transformsants  
 CC harboring a nucleotide sequence comprising a promoter operably  
 CC linked to a mutated TD sequence demonstrate increased levels of  
 CC isoleucine production, and thus provide an improved nutrient source.  
 XX  
 SQ Sequence 502 AA;  
 Query Match 21.2%; Score 368; DB 20; Length 502;  
 Best Local Similarity 30.2%; Pred. No. 2.8e-28;  
 Matches 102; Conservative 65; Mismatches 129; Indels 42; Gaps 10;  
 QY 26 TPVLTSILNQLTGNLFFKCFELQKTSFKIRGALNAVRSILVDPALERKPAVYTHSSG 85  
 Db 21 SPLQLAKLSKRLGVRMYLKRDLQPVFSFKLRGAYNMVKLPADQL---AKGVICSSAG 77  
 QY 86 NHGQALTYAAKLEGIPAYIVPQTAPDCKKLAIQAYGASIVYCEPSDESRENVAKRVTTE 145  
 Db 78 NHAOGVALSASKLGCTAVIVMPVTPEIKQWAVENLGATVVLFGDSYDQAAHAK-IRAE 136  
 QY 146 TEGI-MVHPNQBPVAVIAGOGTIALEVLNQV--PLVDALVVPVGGGMLAGTIAITVKALKP 202  
 Db 137 EELGTFIPFDPHDPVAVIAGOGTIVGMEITRQAKGPL-HAIFVPVGGGGLGIAGIAAYVKRVSP 195  
 QY 203 SVKYVAEAPSNAADCCYQSKLGLKMPNLYPPETI-----ADGVK-SSIGLNTWPIIR 253  
 Db 196 EVKIIGVEPADAN-----AMALSLHHGERVILDQVGGFADGVAVKEVEEETFRISR 246  
 QY 254 DLVDDIFTVTEDEIKCATQLVWERMKLLIEPTAGVGVAAVLQSQHFQTVSPVKNICIVLS 313  
 Db 247 NLMGCVLVTTRDAICASIKDMFEKRNILEPAGALALAGA-EAYCKYYGLKDVNVVAITS 305  
 QY 314 GGNVDL-----TSSITVWVKQAEPPASYQ 336  
 Db 306 GANNMFDKLRIVTELAVNGROQEAVALTMPEKPGSFK 343  
 RESULT 13  
 AAY32943  
 ID AAY32943 standard; Protein; 532 AA.  
 XX  
 AC AAY32943;  
 XX  
 DT 09-NOV-1999 (first entry)  
 XX  
 DE Mutant threonine dehydratase/deaminase protein sequence.  
 XX  
 KW Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; muten.  
 XX  
 OS Arabidopsis thaliana.  
 OS Synthetic.  
 XX  
 PN WO9941395-A1.  
 XX  
 PD 19-AUG-1999.  
 XX  
 PF 08-JAN-1999; 99WO-US00560.  
 XX  
 PR 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX  
 PA (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.

XX Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 PI WPI: 1999-527375/44.  
 XX N-PSDB; AA11201.  
 DR  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX  
 PS Claim 13; Page 111-114; 194pp; English.  
 XX  
 CC This sequence represents a mutant Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
 CC marker (imparting resistance to toxic structural analogues of isoleucine)  
 CC for selecting transformed cells and to produce transformants with  
 CC increased levels of isoleucine (and thus better nutritional value) or of  
 CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for  
 CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
 CC TD-expressing plants permit use of the isoleucine structural analogues as  
 CC herbicides. The DNA sequences are alternatives for antibiotic resistance  
 CC markers (which are potentially harmful to the environment). Since no  
 CC human analog of TD exists (humans can not synthesize isoleucine), it  
 CC should be safe to use.  
 XX  
 SQ Sequence 532 AA;  
 Query Match 21.2%; Score 368; DB 20; Length 532;  
 Best Local Similarity 30.2%; Pred. No. 3.1e-28;  
 Matches 102; Conservative 65; Mismatches 129; Indels 42; Gaps 10;  
 QY 26 TPVLTSILNQLTGNLFFKCFELQKTSFKIRGALNAVRSILVDPALERKPAVYTHSSG 85  
 Db 51 SPLQLAKLSKRLGVRMYLKRDLQPVFSFKLRGAYNMVKLPADQL---AKGVICSSAG 107  
 QY 86 NHGQALTYAAKLEGIPAYIVPQTAPDCKKLAIQAYGASIVYCEPSDESRENVAKRVTTE 145  
 Db 108 NHAOGVALSASKLGCTAVIVMPVTPEIKQWAVENLGATVVLFGDSYDQAAHAK-IRAE 166  
 QY 146 TEGI-MVHPNQBPVAVIAGOGTIALEVLNQV--PLVDALVVPVGGGMLAGTIAITVKALKP 202  
 Db 167 EELGTFIPFDPHDPVAVIAGOGTIVGMEITRQAKGPL-HAIFVPVGGGGLGIAGIAAYVKRVSP 225  
 QY 203 SVKYVAEAPSNAADCCYQSKLGLKMPNLYPPETI-----ADGVK-SSIGLNTWPIIR 253  
 Db 226 EVKIIGVEPADAN-----AMALSLHHGERVILDQVGGFADGVAVKEVEEETFRISR 276  
 QY 254 DLVDDIFTVTEDEIKCATQLVWERMKLLIEPTAGVGVAAVLQSQHFQTVSPVKNICIVLS 313  
 Db 277 NLMGCVLVTTRDAICASIKDMFEKRNILEPAGALALAGA-EAYCKYYGLKDVNVVAITS 335  
 QY 314 GGNVDL-----TSSITVWVKQAEPPASYQ 336  
 Db 336 GANNMFDKLRIVTELAVNGROQEAVALTMPEKPGSFK 373  
 RESULT 14  
 AAY05707  
 ID AAY05707 standard; Protein; 532 AA.  
 XX  
 AC AAY05707;  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE Feedback insensitive mature threonine dehydratase/deaminase.  
 XX  
 KW Threonine dehydratase/deaminase; omr1 gene; feedback inhibition;  
 KW transgenic plant; selectable marker; isoleucine; mutant.  
 XX  
 OS Arabidopsis thaliana.  
 OS Synthetic.  
 XX

FH Key Location/Qualifiers  
 FT Misc-difference 439 /note= "Arg in wild-type enzyme"  
 FT Misc-difference 484 /note= "Arg in wild-type enzyme"  
 FT Region 427..444  
 FT /note= "regulatory region R4"  
 FT Region 476..494  
 FT /note= "regulatory region R6"  
 XX W09902656-Al.  
 PN 21-JAN-1999.  
 XX 10-JUL-1998; 98WO-US14362.  
 XX 17-FEB-1998; 98US-0074875.  
 PR 10-JUL-1997; 97US-0052096.  
 XX (PURD ) PURDUE RES FOUND.  
 XX Mourad GS;  
 XX WPI; 1999-120860/10.  
 DR N-PSDB; AAX25336.  
 XX New sequences encode mutant threonine dehydratase/deaminase - which is insensitive to feedback inhibition, useful as a selective marker to produce transformed cells resistant to toxic isoleucine analogues  
 XX Disclosure; Page 65-68; 120pp; English.  
 XX This sequence represents an Arabidopsis thaliana mutant mature threonine dehydratase/deaminase (TD) which, unlike the wild-type enzyme (see AAX05702), is insensitive to feedback inhibition by isoleucine. It lacks the chloroplast transit peptide of the mutant TD precursor, and is encoded by a polynucleotide (see AAX25336) that is one of 9 claimed polynucleotides (see AAX25332-40), originally isolated and cloned from A. thaliana mutant line GM11b (omrl/omrl), which encode a feedback insensitive TD. These polynucleotides can be used to transform a wide variety of plants, fungi, bacteria and yeast. These polynucleotides differ from the wild-type only by 2 point mutations, which result in R499C and R554H amino acid substitutions (numbering according to wild-type TD) in the R4 and R6 regulatory regions. These forms of TD are not only insensitive to feedback inhibition by isoleucine, but are also insensitive to structural analogues of isoleucine that are toxic to plants and microorganisms which synthesise only wild-type TD. Nucleotide sequences encoding mutated forms of TD can therefore be used to create cells that are insensitive to compounds normally toxic to cells expressing only wild-type TD enzymes, and thus may be used to provide a biochemical selectable marker. Transforms harboring a nucleotide sequence comprising a promoter operably linked to a mutated TD sequence demonstrate increased levels of isoleucine production, and thus provide an improved nutrient source.  
 XX Sequence 532 AA;  
 SQ  
 Query Match 21.2%; Score 368; DB 20; Length 532;  
 Best Local Similarity 30.2%; Pred. No. 3.le-28;  
 Matches 102; Conservative 65; Mismatches 129; Indels 42; Gaps 10;  
 QY 26 TPVLTSILNQLTGRNLFKCELFQKTSFKIRGALNVRSLVPDALEKPKAVVTHSSG 85  
 Db 51 SPLQAKLSKSLRGVRYMLKREDLPVFSFKLRGAYNMVKLPADQL---AKGVTCSSAG 107  
 QY 86 NHGQALTYAAKLEGIPAYIVPQTPADCKKLAIQAGISVYCEPDESRENKVRTEE 145  
 Db 108 NHAQGVALSASKLGCTAVTMVPTVTEIKWQAVENLGATVFLFGSDYDQAQAHAK-IRAE 166  
 QY 146 TESI-MVHPNQPAVIAGOGTIALEVLNQV--PLVDALVVPVGGGMLAGIAITVTKALP 202  
 Db 167 EELGTFIPFPDHPDIAGOGTVGMEITROAKGGL-HAIFVPVGGGGLIAGIAAYKRVSP 225

QY 203 SVKVAAPSNADDCYQSKLGLKMPNLYPPETI-----ADGVK-SSIGLNTWPIIR 253  
 Db 226 EVKIIGVEPADAN-----AMALSLHGERVILDQVGGFADGVAVKEVEETFRISR 276  
 QY 254 DLVDDIFTVTEDEIKATQLVWERMKLLIEPTAGVGAVALSOHFQTVSPVKNICIVLS 313  
 Db 277 NLMGCVLVLTDAICASTKDMFEKRNILEPAGALAGA-EAYCKYGLKDVNVVAITS 335  
 QY 314 GGNVDL-----TSSITWVKQAEPPASQY 336  
 Db 336 GAMNFDKLRIVTELANVGROQEAVALTLMPKPGSEK 373

RESULT 15  
 AAY32942  
 ID AAY32942 standard; Protein; 539 AA.  
 XX  
 AC AAY32942;  
 XX  
 DT 09-NOV-1999 (first entry)  
 XX  
 DE Mutant threonine dehydratase/deaminase protein sequence.  
 XX  
 KW Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; mutein.  
 XX  
 OS Arabidopsis thaliana.  
 OS Synthetic.  
 XX  
 PN W09941395-Al.  
 XX  
 PD 19-AUG-1999.  
 XX  
 PF 08-JAN-1999; 99WO-US00560.  
 XX  
 PR 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX  
 PA (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.  
 XX  
 PI Larrinua IM, Merlo DJ, Mourad GS, Paredy DR;  
 XX  
 DR WPI; 1999-527375/44.  
 DR N-PSDB; AAX11200.  
 XX  
 PT New nucleic acid encoding threonine dehydratase deaminase resistant to feedback inhibition, useful as selection marker for cell transformation and to impart herbicide resistance  
 PT  
 PT  
 XX Claim 13; Page 106-109; 194pp; English.  
 XX  
 CC This sequence represents a mutant Arabidopsis thaliana threonine dehydratase/deaminase (TD) protein of the invention. The protein is a feedback insensitive mutant. The TD DNA sequence is used as molecular marker (imparting resistance to toxic structural analogues of isoleucine) for selecting transformed cells and to produce transformants with increased levels of isoleucine (and thus better nutritional value) or intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also TD-expressing plants permit use of the isoleucine structural analogues as herbicides. The DNA sequences are alternatives for antibiotic resistance markers (which are potentially harmful to the environment). Since no human analog of TD exists (humans can not synthesize isoleucine), it should be safe to use.  
 CC  
 CC Sequence 539 AA;  
 SQ  
 Query Match 21.2%; Score 368; DB 20; Length 539;  
 Best Local Similarity 30.2%; Pred. No. 3.le-28;



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 04:18:31 ; Search time 21.5317 Seconds  
(without alignments)  
464.608 Million cell updates/sec

Title: us-09-889-609B-10  
Perfect score: 1735  
Sequence: 1 MCAQYCISFADVEKAHINIR.....SSITWVKAERPASYQSVSV 340

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Issued\_Patents\_AA:\*
- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*
  - 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*
  - 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*
  - 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*
  - 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*
  - 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1735	100.0	340	US-09-789-300A-2	Sequence 2, Appli
2	367	21.2	424	US-09-134-001C-3876	Sequence 3876, Ap
3	348.5	20.1	441	US-08-403-866-10	Sequence 10, Appl
4	310.5	17.9	436	US-08-669-378-2	Sequence 2, Appli
5	310.5	17.9	436	US-08-669-378-12	Sequence 12, Appl
6	309.5	17.8	436	US-08-669-378-4	Sequence 4, Appli
7	309.5	17.8	436	US-08-669-378-6	Sequence 6, Appli
8	309.5	17.8	436	US-08-669-378-10	Sequence 10, Appl
9	307.5	17.7	436	US-08-669-378-8	Sequence 8, Appli
10	283.5	16.3	378	US-09-789-300A-4	Sequence 4, Appli
11	259.5	15.0	325	US-09-088-435-1	Sequence 1, Appli
12	241.5	13.9	367	US-09-134-001C-4168	Sequence 4168, Ap
13	187	10.8	551	US-08-120-960-2	Sequence 2, Appli
14	187	10.8	551	US-09-347-878-9	Sequence 9, Appli
15	174	10.0	312	US-09-134-001C-3920	Sequence 3920, Ap
16	164.5	9.5	319	US-09-134-001C-3330	Sequence 3330, Ap
17	124.5	7.2	499	US-09-594-193-4	Sequence 4, Appli
18	121.5	7.0	404	US-08-887-534A-34	Sequence 34, Appl
19	120	6.9	497	US-09-594-193-2	Sequence 2, Appli
20	106	6.1	487	US-09-594-193-13	Sequence 13, Appl
21	105	6.1	540	US-08-973-462-22	Sequence 22, Appl
22	105	6.1	1786	US-08-973-462-8	Sequence 8, Appli
23	101.5	5.9	212	US-08-973-462-25	Sequence 25, Appl
24	101.5	5.9	630	US-08-973-462-9	Sequence 9, Appli
25	99.5	5.7	484	US-09-066-046-8	Sequence 8, Appli
26	99	5.7	1642	US-08-662-227-2	Sequence 2, Appli
27	99	5.7	1642	US-09-017-947-2	Sequence 2, Appli

28	98.5	5.7	434	4	US-09-594-193-7	Sequence 7, Appli
29	98	5.6	403	4	US-09-134-001C-5236	Sequence 5236, Ap
30	93	5.4	1648	2	US-08-662-227-35	Sequence 35, Appl
31	93	5.4	1648	4	US-09-017-947-35	Sequence 35, Appl
32	92.5	5.3	619	4	US-09-066-046-2	Sequence 2, Appli
33	91	5.2	1582	2	US-08-404-531B-9	Sequence 9, Appli
34	91	5.2	1582	3	US-08-476-900A-9	Sequence 9, Appli
35	91	5.2	1582	3	US-08-488-546A-9	Sequence 9, Appli
36	91	5.2	1582	4	US-08-726-320-5	Sequence 5, Appli
37	91	5.2	1582	4	US-09-208-716-5	Sequence 5, Appli
38	89	5.1	1642	1	US-08-447-411-45	Sequence 45, Appl
39	88.5	5.1	388	4	US-09-134-001C-4951	Sequence 4951, Ap
40	88	5.1	395	2	US-08-404-531B-3	Sequence 3, Appli
41	88	5.1	395	3	US-08-476-900A-3	Sequence 3, Appli
42	88	5.1	395	3	US-08-488-546A-3	Sequence 3, Appli
43	88	5.1	1580	4	US-08-726-320-1	Sequence 1, Appli
44	88	5.1	1580	4	US-09-208-716-1	Sequence 1, Appli
45	88	5.1	1581	2	US-08-404-531B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-09-789-300A-2  
: Sequence 2, Application US/09789300A  
: Patent No. 6458576  
: GENERAL INFORMATION:  
: APPLICANT: Meyers, Rachel  
: APPLICANT: Rudolph-Owen, Laura A.  
: TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate  
: FILE REFERENCE: 35800/208926  
: CURRENT APPLICATION NUMBER: US/09/789.300A  
: CURRENT FILING DATE: 2001-02-20  
: PRIOR APPLICATION NUMBER: US 60/183,208  
: PRIOR FILING DATE: 2000-02-17  
: NUMBER OF SEQ ID NOS: 4  
: SOFTWARE: FastSeq for Windows Version 4.0  
: SEQ ID NO 2  
: LENGTH: 340  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-09-789-300A-2

Query Match	100.0%	Score 1735;	DB 4;	Length 340;
Best Local Similarity	100.0%	Pred. No. 2.1e-189;	Mismatches 0;	Gaps 0;
Matches 340;	Conservative 0;			
Qy	1	MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLPFFKCELFORTGSFKIRGA	60	
Db	1	MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLPFFKCELFORTGSFKIRGA	60	
Qy	61	LNAVRSVPDALEKRPKAVVTHSSNGHQALTYAAKLEGTPAYIVVPTAPDCKKLAIOA	120	
Db	61	LNAVRSVPDALEKRPKAVVTHSSNGHQALTYAAKLEGTPAYIVVPTAPDCKKLAIOA	120	
Qy	121	YGASIVVCEPDESRENVAKRVTEETEGIMVHPNQEPVIAVQOGTIALEVLNQVPLVDAL	180	
Db	121	YGASIVVCEPDESRENVAKRVTEETEGIMVHPNQEPVIAVQOGTIALEVLNQVPLVDAL	180	
Qy	181	VVPVGGGMLAGIAITVKALKPSVKVYAAEPNADDCYQSKLKGKLPNLYPPETIADGV	240	
Db	181	VVPVGGGMLAGIAITVKALKPSVKVYAAEPNADDCYQSKLKGKLPNLYPPETIADGV	240	
Qy	241	KSSIGLNTWPIIRDLDVDDIFVTVEDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFQT	300	
Db	241	KSSIGLNTWPIIRDLDVDDIFVTVEDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFQT	300	
Qy	301	VSPEVKNICIVLSCGNVDLTSSITWVKAERPASYQSVSV	340	
Db	301	VSPEVKNICIVLSCGNVDLTSSITWVKAERPASYQSVSV	340	

NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30, 727  
REFERENCE/DOCKET NUMBER: 20747/30  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1600  
TELEFAX: (716) 263-1487  
TELEX: 978450 (WUT)  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 441 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Lactococcus lactis subsp. lactis  
INDIVIDUAL ISOLATE: ILVA  
US-08-403-866-10

Query Match	20.1%	Score 348.5;	DB 1;	Length 441;
Best Local Similarity	29.9%;	Pred. No. 6.7e-31;		
Matches	95;	Conservative 65;	Mismatches 137;	Indels 21; Gaps 8;

  

QY	14	KAHINIRDSHTLPVLTSSLTNLGTGRNLFKCELFQKTSFKIRGALNAVRSIVPDALE	73
DB	34	QANIYLKEVTKTPLODPYLSNKYQANIYLKEENLQKVRSLKRGAYYSISKL---	90
QY	74	RPKPAAVTHSSNGHQALTYAAKLEGIPAYIVVPQTAPDCKLIAQAYGASIVVC---	129
DB	91	QRSGKVVASAGNHAOGVAFAANQLNLSATIFMPVTPNQKISQVRFEGESHVTRILIGD	150
QY	130	PSDESRENVAKRVTEETEGIMVHPNGEPAVIAGQGTIALEVLNQVP----	185
DB	151	TFDES--ARAAKAFSQNDNPPFIDPFDDENVIAGQGTVALEIFAQAQKQGISLQKIFVQIG	209
QY	186	GGGMLAGIATVYKALXPSVKVYAAEAPSNDCCYQSKLKGKMLPNLPPEIADGVK--	244
DB	210	GGGLIAGITAYSKERYPQTEIIGVEAKGATSMKAAYSAGQPV--TLEHIDKFAQGIATV	268
QY	245	GLMTWPIIRDLVDDIFTVTEDEIKCATQLVWERSMKLLIETAGVGVAAVLSQHFQTVSP	304
DB	269	GQKTYOLINDKVKQLLAVDEGLLSQITILELYSKLGIVAEPAGATSVAAAL-----	323
QY	305	V--KNTCIVLSSGNGVDLT	320
DB	324	IKGNITVCIISGNGNDIS	341

RESULT 4  
 US-08-669-378-2  
 Sequence 2, Application US/08669378  
 Patent No. 6107063  
 GENERAL INFORMATION:  
 APPLICANT: Moeckel, Bettina  
 APPLICANT: Eggeling, Lothar  
 APPLICANT: Sahm, Hermann  
 TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF  
 TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE  
 TITLE OF INVENTION: DEHYDRATASE  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W., Suite 500  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/669,378  
FILING DATE: 20-MAR-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/DE95/00017  
FILING DATE: 09-JAN-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 00 926.7  
FILING DATE: 14-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 016881/0142  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-669-378-2

Query Match 17.9%; Score 310.5; DB 3; Length 436;  
Best Local Similarity 27.8%; Pred. No. 1.4e-26;  
Matches 93; Conservative 61; Mismatches 160; Indels 21; Gaps 9;  
Qy 7 ISFADVEKAHINIRDSIHLTPVLTSSILNLTGRNLEFFKCELFQKTSFKIRGALNAVRS 66  
Db 21 IRAADIQAARISSVIAPTLPQYCPRLSEETGAETIYKREDLQDVRSYKIRGALN---S 77  
Qy 67 LVPDALEKPKAVVTHSSGNHGALTYAAKLEGIPAVIVPQAPDCCKKLAIOAYG--A 123  
Db 78 GAOSPOQRDAGIVAASAGNHAOGVAVCKSLGVQGRYIYVPQPKOKRDRINWVGFEV 137  
Qy 124 SIYCEPDESRENKRVTEETEGIMVHPNQEPVIAIGOGTIALEVLNQVPLV----DA 179  
Db 138 SLVVTGNFDEASAAAHEDAERTGATLIEPFDARNTVIGQGTVAEILSOLTSNGKSADH 197  
Qy 180 LVVPVGGGMLAGIATVVKALKPSVYAAEPSNADDCYOSKLGKLMPLNYPPEIADG 239  
Db 198 VMVPVGGGGLLAGVSYVMADMAPRTAIVGIEPAGAAS--MQAALHNGGPIITLETVDPEVDG 256  
Qy 240 VK--SSIGLNTWPIIRDLVDIFVTDEIKCATOL--VWERMKLLIEPTAGVGVAAVLSQ 296  
Db 257 AEVQRVGDLYNTIVEKNQGRVHMMSATEGAVCTEMLDLYONEGLIAEPAGALSIAGLKEM 316  
317 SF---APGSVVVCII-SGGNDV---LRYAEIAER 344

RESULT 5  
US-08-669-378-12  
Sequence 12, Application US/08669378  
Patent No. 6107063  
GENERAL INFORMATION:  
APPLICANT: Moeckel, Bettina  
APPLICANT: Eggeling, Lothar  
APPLICANT: Sahm, Hermann  
TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF  
TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGLATED THREONINE  
TITLE OF INVENTION: DEHYDRATASE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/669,378  
FILING DATE: 20-MAR-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/DE95/00017  
FILING DATE: 09-JAN-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 00 926.7  
FILING DATE: 14-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 016881/0142  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-669-378-12

Query Match 17.9%; Score 310.5; DB 3; Length 436;  
Best Local Similarity 28.1%; Pred. No. 1.4e-26;  
Matches 94; Conservative 60; Mismatches 160; Indels 21; Gaps 9;  
Qy 7 ISFADVEKAHINIRDSIHLTPVLTSSILNLTGRNLEFFKCELFQKTSFKIRGALNAVRS 66  
Db 21 IRAADIQAARISSVIAPTLPQYCPRLSEETGAETIYKREDLQDVRSYKIRGALN---S 77  
Qy 67 LVPDALEKPKAVVTHSSGNHGALTYAAKLEGIPAVIVPQAPDCCKKLAIOAYG--A 123  
Db 78 GAOSPOQRDAGIVAASAGNHAOGVAVCKSLGVQGRYIYVPQPKOKRDRINWVGFEV 137  
Qy 124 SIYCEPDESRENKRVTEETEGIMVHPNQEPVIAIGOGTIALEVLNQVPLV----DA 179  
Db 138 SLVVTGNFDEASAAAHEDAERTGATLIEPFDARNTVIGQGTVAEILSOLTSNGKSADH 197  
Qy 180 LVVPVGGGMLAGIATVVKALKPSVYAAEPSNADDCYOSKLGKLMPLNYPPEIADG 239  
Db 198 VMVPVGGGGLLAGVSYVMADMAPRTAIVGIEPAGAAS--MQAALHNGGPIITLETVDPEVDG 256  
Qy 240 VK--SSIGLNTWPIIRDLVDIFVTDEIKCATOL--VWERMKLLIEPTAGVGVAAVLSQ 296  
Db 257 AEVQRVGDLYNTIVEKNQGRVHMMSATEGAVCTEMLDLYONEGLIAEPAGALSIAGLKEM 316  
317 SF---APGSVVVCII-SGGNDV---LRYAEIAER 344

RESULT 6  
US-08-669-378-4  
Sequence 4, Application US/08669378  
Patent No. 6107063  
GENERAL INFORMATION:  
APPLICANT: Moeckel, Bettina  
APPLICANT: Eggeling, Lothar  
APPLICANT: Sahm, Hermann  
TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF  
TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGLATED THREONINE  
TITLE OF INVENTION: DEHYDRATASE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 20-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DE95/00017  
FILING DATE: 09-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 00 926.7  
FILING DATE: 14-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 016881/0142  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-669-378-4

Query Match 17.8%; Score 309.5; DB 3; Length 436;  
Best Local Similarity 27.8%; Pred. No. 1.9e-26;  
Matches 93; Conservative 61; Mismatches 160; Indels 21; Gaps 9;  
QY 7 ISPADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVR 66  
Db 21 IRAADIQTAAQTSSVIPTPLQYCPRLSEETGAEIYLKREDLQDVRSYKIRGALN---S 77  
QY 67 LVPDALKERPKAVVTHSSNGHQAITYAAKLEGPAYIVVQTPADCKKLAIOAYG---A 123  
Db 78 GAOSPOQRDAGIVAASAGNHAQGVAYVCKSLGVQGRYIVPVQTPKQKDRIMVHGGEFV 137  
QY 124 SIYVCEPSDESRENKRVTEETEGIMVHPNQEPVAVIAGOGTIALEVLNOVPLV---DA 179  
Db 138 SLVVTGNFDEASAAHEDARTGATLIEPFDARNVTIGOGTVAAILSLTSMGKSADH 197  
QY 180 LVVPVGGGMLAGIATTVKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNLYPETIADG 239  
Db 198 VMVPVGGGGLLAGVSYVMADMAPRTAIVGIEPAGAAS-MQAALHNGGPITLETVDVDFVDG 256  
QY 240 VK--SSIGLNTWPIIRDLVDDIFVTEDFKCATQL--VWERMKLLIETPAGVGVAAVLSQ 296  
Db 257 AEVKRYGDLNLTIVKQGRVHMSATGAVCTEMLDLYQNEGIIAEPAGALSIALGKEM 316  
QY 297 HFQTVSPVKNICIVLSSGGNVDLTSSITWVKQAE 331  
Db 317 SF---APGSVVVCI--SGGNNDV---LRYAEIAER 344

## RESULT 7

US-08-669-378-6  
Sequence 6, Application US/08669378.  
Patent No. 6107063  
GENERAL INFORMATION:  
APPLICANT: Moeckel, Bettina  
APPLICANT: Eggeling, Lothar

APPLICANT: Sahn, Hermann  
TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF  
TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE  
TITLE OF INVENTION: DEHYDRATASE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 20-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DE95/00017  
FILING DATE: 09-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 00 926.7  
FILING DATE: 14-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 016881/0142  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-669-378-6

Query Match 17.8%; Score 309.5; DB 3; Length 436;  
Best Local Similarity 27.8%; Pred. No. 1.9e-26;  
Matches 93; Conservative 61; Mismatches 160; Indels 21; Gaps 9;  
QY 7 ISPADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVR 66  
Db 21 IRAADIQTAAQTSSVIPTPLQYCPRLSEETGAEIYLKREDLQDVRSYKIRGALN---S 77  
QY 67 LVPDALKERPKAVVTHSSNGHQAITYAAKLEGPAYIVVQTPADCKKLAIOAYG---A 123  
Db 78 GAOSPOQRDAGIVAASAGNHAQGVAYVCKSLGVQGRYIVPVQTPKQKDRIMVHGGEFV 137  
QY 124 SIYVCEPSDESRENKRVTEETEGIMVHPNQEPVAVIAGOGTIALEVLNOVPLV---DA 179  
Db 138 SLVVTGNFDEASAAHEDARTGATLIEPFDARNVTIGOGTVAAILSLTSMGKSADH 197  
QY 180 LVVPVGGGMLAGIATTVKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNLYPETIADG 239  
Db 198 VMVPVGGGGLLAGVSYVMADMAPRTAIVGIEPAGAAS-MQAALHNGGPITLETVDVDFVDG 256  
QY 240 VK--SSIGLNTWPIIRDLVDDIFVTEDFKCATQL--VWERMKLLIETPAGVGVAAVLSQ 296  
Db 257 AEVKRYGDLNLTIVKQGRVHMSATGAVCTEMLDLYQNEGIIAEPAGALSIALGKEM 316  
QY 297 HFQTVSPVKNICIVLSSGGNVDLTSSITWVKQAE 331  
Db 317 SF---APGSVVVCI--SGGNNDV---LRYAEIAER 344

## RESULT 8

```
US-08-669-378-10
: Sequence 10, Application US/08669378
: Patent No. 6107063
: GENERAL INFORMATION:
: APPLICANT: Moeckel, Bettina
: APPLICANT: Eggeling, Lothar
: APPLICANT: Sahm, Hermann
: TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
: TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
: TITLE OF INVENTION: DEHYDRATASE
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/669,378
: FILING DATE: 20-MAR-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION NUMBER: PCT/DE95/00017
: FILING DATE: 09-JAN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE P 44 00 926.7
: FILING DATE: 14-JAN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Bent, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 016881/0142
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 436 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-669-378-10

Query Match . 17.8%; Score 309.5; DB 3; Length 436;
Best Local Similarity 27.8%; Pred. No. 1.9e-26;
Matches 93; Conservative 61; Mismatches 160; Indels 21; Gaps 9;

QY 7 ISFADVEKAHINTRSIHLPVLTSSILNQLTGRNLFKCELPQKGTGSKIRGALNAVRS 66
Db 21 IRAADIQTAAQARISSVIAPTPLQYCPRLSEETGAETIYLKREDLQDVRSYKIRGALN---S 77
QY 67 LVPDALERPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTQAPDCKKLAIOAYG---A 123
Db 78 GAQSPQEQDAGIVAASAGNHAOGVAYVCKSLGVQGRIVVPTQPKOKRDRIMVHGFEV 137
QY 124 SIYCEPSPDESRENKRVTEETEGIMVHPNOEPVIAQGQTTALEVLNQVPLV----DA 179
Db 138 SLVVTGNFDEASAAHEDAERTGATLIEPFDARNVIGOGTVAAILSOLTSMGKSADH 197
QY 180 LVVPVGGGMLAGIATTVKALPKSVKYAAEPSNADDCYOSKLGKGLMPLNPPETIADG 239
Db 198 VWVPVGGGGLAGVSWYMDMAPRTAIVGTETEPAGAAS--MQAALHNGGPITLTETVDFVDG 256
QY 240 VK--SSIGLNTWPIIRDVDDIFVTDEIKCATOL--VWERMKLLIETAGVCAVAVLSQ 296
Db 257 AEVKRVGDLNYYTIVEKNQGRVHMMSATEGAVCTEMLDLYONEGIIAEPAULSIAGLKEM 316

US-08-669-378-8
: Sequence 8, Application US/08669378
: Patent No. 6107063
: GENERAL INFORMATION:
: APPLICANT: Moeckel, Bettina
: APPLICANT: Eggeling, Lothar
: APPLICANT: Sahm, Hermann
: TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
: TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
: TITLE OF INVENTION: DEHYDRATASE
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/669,378
: FILING DATE: 20-MAR-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION NUMBER: PCT/DE95/00017
: FILING DATE: 09-JAN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE P 44 00 926.7
: FILING DATE: 14-JAN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Bent, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 016881/0142
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 436 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-669-378-8

Query Match . 17.7%; Score 307.5; DB 3; Length 436;
Best Local Similarity 28.6%; Pred. No. 3.1e-26;
Matches 96; Conservative 59; Mismatches 158; Indels 23; Gaps 10;

QY 7 ISFADVEKAHINTRSIHLPVLTSSILNQLTGRNLFKCELPQKGTGSKIRGALNAVRS 66
Db 21 IRAADIQTAAQARISSVIAPTPLQYCPRLSEETGAETIYLKREDLQDVRSYKIRGALN---S 77
QY 67 LVPDALERPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTQAPDCKKLAIOAYG---A 123
Db 78 GAQSPQEQDAGIVAASAGNHAOGVAYVCKSLGVQGRIVVPTQPKOKRDRIMVHGFEV 137
QY 124 SIYCEPSPDESRENKRVTEETEGIMVHPNOEPVIAQGQTTALEVLNQVPLV----DA 179
Db 138 SLVVTGNFDEASAAHEDAERTGATLIEPFDARNVIGOGTVAAILSOLTSMGKSADH 197
QY 180 LVVPVGGGMLAGIATTVKALPKSVKYAAEPSNADDCYOSKLGKGLMPLNPPETIADG 239
Db 198 VWVPVGGGGLAGVSWYMDMAPRTAIVGTETEPAGAAS--MQAALHNGGPITLTETVDFVDG 256
QY 240 VK--SSIGLNTWPIIRDVDDIFVTDEIKCATOL--VWERMKLLIETAGVCAVAVLSQ 296
Db 257 AEVKRVGDLNYYTIVEKNQGRVHMMSATEGAVCTEMLDLYONEGIIAEPAULSIAGLKEM 316
```

Db 198 VMPVGGGGLLAGVWSYMDAMPRTAIVGIERPAGAAS-MQAALHNGSPITLETVPFVDG 256  
QY 240 --VKSSIGLNTWPIIRDVLVDIFVTDEIKCATQL--VWERMKLIEPTAGVGVAAVLS 295  
Db 257 GEYKRVCDLN-YTIVEKNQGRVHMSTEGAVCTEMLDLYQNEGIIAEPAGALSIAGLKE 315  
QY 296 QHFTQVSPVKNKICIVLSGGNVDLTSSITVWQKAR 331  
Db 316 MSF---APGSVWVCII-SGGNDV---LRYAETAER 344

RESULT 10  
US-09-789-300A-4  
; Sequence 4, Application US/09789300A  
; Patent No. 6458576  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; APPLICANT: Rudolph-Owen, Laura A.  
; TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate  
; TITLE OF INVENTION: Dependent Enzyme Family Member and Uses Therefore  
; FILE REFERENCE: 35800/208926  
; CURRENT APPLICATION NUMBER: US/09/789,300A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: US 60/183,208  
; PRIOR FILING DATE: 2000-02-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 378  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Pyridoxal-Phosphate Dependent Enzyme Family Domain  
; OTHER INFORMATION: Sequence  
US-09-789-300A-4

Query Match 16.3%; Score 283.5; DB 4; Length 378;  
Best Local Similarity 28.3%; Pred. No. 1.4e-23;  
Matches 109; Conservative 55; Mismatches 126; Indels 95; Gaps 15;

QY 19 IRDSIHLPVLTSSILNQLTGRNL-----PFKCE-LFQKTSKIRG-ALNAV---- 64  
Db 1 VTELGNTPLVRNLRUSKELGEGLGANAAVEIYKLEDLNGPTGSPKDRGLALNMLLAE 60  
QY 65 -----RSLVPDALERPK-AVTHSSGNHQALTYAAKLEGIPAYIVWPOTAPDCKKLA- 117  
Db 61 KLGGKGGIVPGTVQVESKTIIEPTSGNTGIALALAAALLGLKCTIVPATDTSREKRAQ 120  
QY 118 IQAYGASIVCEPSDESRENKRVTEETEGIMVHPNQBPV-----TA 161  
Db 121 LRALGAELVVPVAGGSDDLADAIKAE--LAEEPENAYLLNQAAGPFONPANPEIA 178  
QY 162 GQCTIALELVNV-----PLVDALVVPVGGGMLAGIATVVKALPS-----V 204  
Db 179 GQKTIGPEIWEQGGKEISLGRLPDPAVAPVGGGGTITGIARYLKLNDPGKIDVLELVP 238  
QY 205 KYAAPSNAADDQCYQSKLKGKLMPLNLY-----PPTIADGVKSSIG 245  
Db 239 KVIGVEPGS-----AVLSGLKATITLAGKPLHGRDSKYLLODEPVTLPETKSIGIG 293  
QY 246 LN-----TWPIIRD-----VDDIFTVTEDEIKCATQLVWERMKLLIEPTAGVAA 292  
Db 294 LGVPRVGEFVPPILDELLDRRQIDRVTVTDEEALEARLLAREGILVGPSSGAAVA 353  
QY 293 VLSQHPQTVSP--EVKNICIVLSGG 315  
Db 354 ALKLAKEGKPLNKGKTIIVILSGG 378

RESULT 11  
US-09-088-435-1  
; Sequence 1, Application US/09088435  
; Patent No. 6277619

; GENERAL INFORMATION:  
; APPLICANT: LAL, PREETI  
; APPLICANT: CORLEY, NEIL C.  
; APPLICANT: GUEGLER, KARL J.  
; APPLICANT: PATTERSON, CHANDRA  
; TITLE OF INVENTION: SERINE DEHYDRATASE HOMOLOG  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/088,435  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carrone, Michael C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0512 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 325 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: THPIAZS08  
; CLONE: 2752518  
; US-09-088-435-1

Query Match 15.0%; Score 259.5; DB 4; Length 325;  
Best Local Similarity 28.6%; Pred. No. 5.8e-21;  
Matches 97; Conservative 58; Mismatches 113; Indels 71; Gaps 16;

QY 10 ADVERAHINIRDSIHLPVLTSSILNQLTGRNLFFKCELFQKTSKIRGALNAVRLVP 69  
Db 9 AKQEPFHV-----VTPLESWALSQVAGMPVLKCNVQPSGSKIRG-----IGHFCQ 57  
QY 70 DALEKRPKAVVTHSSGNHQALTYAAKLEGIPAYIVWPOTAPDCKKLAIQAYGAST---- 125  
Db 58 EMNAKGRHLVCGSSGNAGIAAAYAAKLGIPATIVLPSTSLQVVQRLQEGEAEVQLTG 117  
QY 126 -VYCEPSDESRENKRVTEETEGIMVHPNQBPVIAQGTIALE---VLNQVPLVDALV 181  
Db 118 KYWDEANLRAQE-LAKRDGWN-----VPPFDHPLHWKSHASLVQELKAVLRTPP--GALV 170  
QY 182 VPWGGGMLAGIATVVKAL-----KPSVKVYAAEPSNADDQCYQSKL-KGKLMPLNLYPETI 236  
Db 171 LAVGGGGLLAGVAGLVLEGVQHOVPIIAHGAH-----CFNAAITAGKLV-----TL 217  
QY 237 AD--GVKSSIGLNT-----WPIIRDVLVDIFVTDEIKCATQLVWERMKLLIE 283  
Db 218 PDITSVAKSLGAKTVAARALECMQVCKTHSEVVEDTEAVS-----AVQQLDDERMLVE 271  
QY 284 PTAGVGVAAVLS-----QHFTQVSPVKNKICIVLSGGN 316  
Db 272 PAC--GAAAIYSGLLRRLRQAEGLPPLPSLTSSVVVIVCGGN 308

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/120,960
; FILING DATE: 12-SEP-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: PEPPER PH.D., FREDERICK W.
; REGISTRATION NUMBER: 31,286
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4410
; TELEFAX: 619-453-2839
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-120-960-2

Query Match 10.8%; Score 187; DB 1; Length 551;
Best Local Similarity 25.2%; Pred. No. 2.7e-12;
Matches 86; Conservative 46; Mismatches 149; Indels 60; Gaps 11;

QY 26 TPVLTSSILNQLTGR--NLFFKCELFQKTSFKIRGALNAVRSVLPPDALERKPKAVVTHS 83
DB 87 TPVNRINKIGKFKGLKCELLAKCEFFNAGGSVKDRISLRMTIEDAERDGTLPKPGDTIIEPT 146
QY 84 SGNHGQALTYAAKLEGIPATVVPQADPCGKKLAIQAYGASIVYCE-----PSDESRENV 138
DB 147 SGTGTGICLAAAVRGYRCILIVPEKMSSEKVDVLRALGAEIFVTPTNARFDSFESHVGV 206
QY 139 AKRVTEE--TEGIMVHPNQBPVITAGGTTALEVLNQVP-LVDALVVPVGGGMLAGTAT 195
DB 207 AWRLKNEIPNSHILDQYRNASNPLAHYDITTADELQCCDGLDMLVASVGTGGTITGIAR 266
QY 196 TVRKALPSKVYAAEPSNADDCYQSKLGLKMLPNLYPETIADGVKSS-----ICLNTWP 250
DB 267 KLEKCPGCRILIGVDP-----EGSI---LAEPEELNQTEQTYEVEGIGYDFIP 312
QY 251 II--RDLDVDDIFTVTEDEIKCATOLVWERMKLLIEPTAG--VGVAANVLSQHFTQVSPVK 306
DB 313 TVLDRITVVDKFRSNDDEATTFARMLTAQEGLLCGSGAGSIVAVAKAAQELQE-----G 367
QY 307 NICIVLSGGNV-----DLTSSITW 325
DB 368 QRCVWILPDSVRNVTMFKFLSDRWMLQKGLKEEDLTEKKPW 408

RESULT 14
US-09-347-878-9
; Sequence 9, Application US/09347878C
; Patent No. 6376210
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
; FILE REFERENCE: 25885-1651
; CURRENT APPLICATION NUMBER: US/09/347,878C
; CURRENT FILING DATE: 1999-07-06
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Homo sapiens
;
US-09-347-878-9

Query Match 10.8%; Score 187; DB 4; Length 551;
Best Local Similarity 25.2%; Pred. No. 2.7e-12;
Matches 86; Conservative 46; Mismatches 149; Indels 60; Gaps 11;

QY 26 TPVLTSSILNQLTGR--NLFFKCELFQKTSFKIRGALNAVRSVLPPDALERKPKAVVTHS 83
DB 87 TPVNRINKIGKFKGLKCELLAKCEFFNAGGSVKDRISLRMTIEDAERDGTLPKPGDTIIEPT 146

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:11:01 ; Search time 276.907 Seconds  
(without alignments)  
132.862 Million cell updates/sec

Title: us-09-889-609B-10  
Perfect score: 1735  
Sequence: 1 MCAQYICISFADVEKAHINR.....SSITWVQAFRPSYQSVS 340

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	Published_Applications_AA:*
1:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2:	/cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	349.5	20.1	310	9	US-09-738-626-4591 Sequence 4591, Ap
2	311.5	18.0	436	9	US-09-738-626-5828 Sequence 5828, Ap
3	192.5	11.1	306	10	US-09-815-242-13657 Sequence 13657, A
4	190.5	11.0	308	9	US-09-738-626-6317 Sequence 6317, Ap
5	174.5	10.1	310	10	US-09-815-242-10707 Sequence 10707, A
6	169.5	9.8	313	10	US-09-815-242-12146 Sequence 12146, A
7	169.5	9.8	313	10	US-09-815-242-12800 Sequence 12800, A
8	169.5	9.8	323	10	US-09-815-242-10239 Sequence 10239, A
9	167.5	9.7	303	10	US-09-815-242-5437 Sequence 5437, Ap
10	161.5	9.3	323	10	US-09-815-242-13862 Sequence 13862, A
11	154	8.9	383	9	US-09-931-457A-65 Sequence 65, Appl
12	149	8.6	324	10	US-09-815-242-5737 Sequence 5737, Ap
13	148	8.5	325	9	US-09-931-457A-32 Sequence 32, Appl
14	147	8.5	324	9	US-09-905-290A-4 Sequence 4, Appli
15	147	8.5	386	9	US-09-931-457A-66 Sequence 66, Appl
16	147	8.5	398	9	US-09-931-457A-62 Sequence 62, Appl
17	146	8.4	415	9	US-09-931-457A-64 Sequence 64, Appl
18	143.5	8.3	326	10	US-09-815-242-11179 Sequence 11179, A
19	143.5	8.3	324	10	US-09-815-242-11879 Sequence 11879, A

20	139	8.0	325	9	US-09-931-457A-31 Sequence 31, Appl
21	131	7.6	306	10	US-09-815-242-11292 Sequence 11292, A
22	120	6.9	317	9	US-09-738-626-5838 Sequence 5838, Ap
23	115.5	6.7	250	10	US-09-815-242-11635 Sequence 11635, A
24	112	6.5	1187	10	US-09-935-291A-4 Sequence 4, Appli
25	105	6.1	540	9	US-09-742-096-22 Sequence 22, Appli
26	105	6.1	1786	9	US-09-742-096-3 Sequence 3, Appli
27	101.5	5.9	212	9	US-09-742-096-25 Sequence 25, Appli
28	101.5	5.9	630	9	US-09-742-096-5 Sequence 5, Appli
29	99	5.7	1642	10	US-09-925-442-2 Sequence 2, Appli
30	96.5	5.6	26926	9	US-09-759-508B-2 Sequence 2, Appli
31	95.5	5.5	472	10	US-09-745-763-36 Sequence 36, Appl
32	95.5	5.5	473	9	US-09-984-271-120 Sequence 120, App
33	95	5.5	164	10	US-09-734-017A-28 Sequence 28, Appl
34	95	5.5	577	10	US-09-815-242-13870 Sequence 13870, A
35	93.5	5.4	852	10	US-09-815-242-11535 Sequence 11535, A
36	93	5.4	577	10	US-09-741-669-354 Sequence 354, App
37	93	5.4	577	10	US-09-815-242-10286 Sequence 10286, A
38	93	5.4	1648	10	US-09-925-442-35 Sequence 35, Appl
39	91.5	5.3	833	9	US-09-928-457-40 Sequence 40, Appl
40	90.5	5.2	442	9	US-10-117-417-6 Sequence 6, Appli
41	90.5	5.2	442	9	US-10-117-417-16 Sequence 16, Appli
42	89.5	5.2	853	10	US-09-815-242-11363 Sequence 11363, A
43	89	5.1	815	9	US-10-198-070-123 Sequence 123, App
44	87.5	5.0	302	9	US-09-738-626-6862 Sequence 6862, Ap
45	87.5	5.0	319	10	US-09-815-242-13897 Sequence 13897, A

ALIGNMENTS

RESULT 1

US-09-738-626-4591  
; Sequence 4591, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAVASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/097738.626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4591  
; LENGTH: 310  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4591

Query Match 20.1%; Score 349.5; DB 9; Length 310;  
Best Local Similarity 29.1%; Pred. No. 8.2e-23;  
Matches 95; Conservative 55; Mismatches 146; Indels 31; Gaps 7;

Qy	7	ISFADVEKAHINRDSIHLTPVLTSIIINLTGNLFKCELFKGTGSKTRGALNAVR	66
Db	2	LTNDVITAOQRAPVHVRTPLEFA---DPIDGQIWIKAETLQKCGVFKTRGAFNQLA	58
Qy	67	LVPDALEKPKAVVTVHSSNGHQAITYAAKLEGIPAYIVVPTQAPDCCKLAIQAYGASIV	126





```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12146
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12146
```

```
Query Match          9.8%; Score 169.5; DB 10; Length 313;
Best Local Similarity 24.5%; Pred. No. 8.4e-08;
Matches 74; Conservative 50; Mismatches 143; Indels 35; Gaps 10;

QY 9 FADVEKAHINIRDSHTLTPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRSVL 68
Db 2 FLMAQKPVNDITQIIIGTTPVVKLRNVDDNAAADVVKLEVQNPGGSVKORIALAMTEKAE 61

QY 69 PDALERKPKAVYTHSSGNHGOALTYAAKLEGIPAYIVVPTAPDCKKKLAIQAYGASIVYC 128
Db 62 REGKIRPGDTIVEPTSGNTGIGLAFVCAAKGYKAVFTMPETMSOERNLLKAYGAELV-L 120

QY 129 EPSDESRENVAKRVTE--ETEGIMVHPNQ-----EPAVIAGOGTTALEVLNV--PLVDA 179
Db 121 TPGSEAMKAIIKAKELKEEHGYF-EPQOFENPANPEV--HELTGPELLOQFEGKTIDA 177

QY 180 LVVPVGGGMLAGIATVVKALPKSVKYAAEPS-----NADDCYOSKLGKMLPNLYPPE 234
Db 178 FLAGVGTGCTLSGVKLVKKEYPNTEIVAIEPEASVLSGGEPGPHKLOG--LGAGFIPG 235

QY 235 TIADGVKSSIGLNTWPIIRDLDVDDIFTVTDEIKCATQLVWERMKLLIPTAGVGVAAVL 294
Db 236 T-----LNT-----EYDSIIKVGNDTAMEMSRVAKREEGILAGISSGAAIYAAI 280

QY 295 SQ 296
Db 281 QK 282
```

```
RESULT 7
US-09-815-242-12800
; Sequence 12800, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12800
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12800
```

```
Query Match          9.8%; Score 169.5; DB 10; Length 313;
Best Local Similarity 24.5%; Pred. No. 8.4e-08;
Matches 74; Conservative 50; Mismatches 143; Indels 35; Gaps 10;

QY 9 FADVEKAHINIRDSHTLTPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRSVL 68
Db 2 FLMAQKPVNDITQIIIGTTPVVKLRNVDDNAAADVVKLEVQNPGGSVKORIALAMTEKAE 61

QY 69 PDALERKPKAVYTHSSGNHGOALTYAAKLEGIPAYIVVPTAPDCKKKLAIQAYGASIVYC 128
Db 62 REGKIRPGDTIVEPTSGNTGIGLAFVCAAKGYKAVFTMPETMSOERNLLKAYGAELV-L 120

QY 129 EPSDESRENVAKRVTE--ETEGIMVHPNQ-----EPAVIAGOGTTALEVLNV--PLVDA 179
Db 121 TPGSEAMKAIIKAKELKEEHGYF-EPQOFENPANPEV--HELTGPELLOQFEGKTIDA 177

QY 180 LVVPVGGGMLAGIATVVKALPKSVKYAAEPS-----NADDCYOSKLGKMLPNLYPPE 234
Db 178 FLAGVGTGCTLSGVKLVKKEYPNTEIVAIEPEASVLSGGEPGPHKLOG--LGAGFIPG 235

QY 235 TIADGVKSSIGLNTWPIIRDLDVDDIFTVTDEIKCATQLVWERMKLLIPTAGVGVAAVL 294
Db 236 T-----LNT-----EYDSIIKVGNDTAMEMSRVAKREEGILAGISSGAAIYAAI 280

QY 295 SQ 296
Db 281 QK 282
```

```
RESULT 8
US-09-815-242-10239
; Sequence 10239, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
```

Query Match	9.7%;	Score 167.5;	DB 10;	Length 303;
Best Local Similarity	24.6%;	Pred. No. 1.2e-07;		
Matches	72:	Conservative	49:	Mismatches 137.
				Indels 35.

Qy	18	NIRDSIHLTPVLVTS	SSILNOLTGRNLFF	CKELFQKTSFKIR	GALNAVRS	SLVDPDALGR	KKPK	77	
Db	2	NITQIGCTPVVKUR	NVDDNAADVVY	KLQYQPGGSV	KDRIALAMIE	KAEREGK	1KPGD	61	
Qy	78	AVVTHSSNHGOAL	TYAAKLEGIPAY	IVVPOPTAPDC	KKLAIQAYGAS	IVYVCEPS	DESREN	137	
Db	62	TIVPTSGNTGIGL	AFVCAAKGYKAV	FTWPTMSQERR	NLLKAYGAELV	-LTPGSEAMK	120		
Qy	138	VAKRVTE--ETEG	IMVHPNQ-----	EPAVIACQGT	IALEVLNOV--	PLVDALV	VPVGGG	188	
Db	121	AIKKAKELKEEH	GYF-EPOQFEN	PANPEYV--	HELTTPG	ELLOQFEG	KTIDAF	LAGVCTGG	177
Qy	189	MLAGTAITVKA	LKPSKVYVAE	PS-----	NADDCYOS	KLKGLM	PNLYP	PETIADGVKSS	243
Db	178	TLSGVGVKLKE	YPNIEIVALE	PEASPLVSG	GEPPHKIQG--	LGAGFI	PGT-----	227	
Qy	244	IGLNTWPIIRDL	VDDIFTVTEDE	IKCATQLVM	RMKLLI	TEPTAGV	GAVALSQ	296	
Db	228	--LNT-----	ETYSIIKVGND	TAMEMSR	RVAKKEG	ILAGISS	GAIAV	AAIOK	273

```

RESULT 10
US-09-815-242-13862
; Sequence 13862, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13862
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Salmonella typhi

```

```

> APPLICANT: Zyskind, Judith W.
> APPLICANT: Wall, Daniel
> APPLICANT: Trawick, John D.
> APPLICANT: Carr, Grant J.
> APPLICANT: Yamamoto, Robert T.
> APPLICANT: Xu, H. Howard
> TITLE OF INVENTION: Identification of Essential Genes in
> TITLE OF INVENTION: Prokaryotes
> FILE REFERENCE: ELITRA.011A
> CURRENT APPLICATION NUMBER: US/09/815,242
> CURRENT FILING DATE: 2001-03-21
> PRIOR APPLICATION NUMBER: 60/191,078
> PRIOR FILING DATE: 2000-03-21
> PRIOR APPLICATION NUMBER: 60/206,848
> PRIOR FILING DATE: 2000-05-23
> PRIOR APPLICATION NUMBER: 60/207,727
> PRIOR FILING DATE: 2000-05-26
> PRIOR APPLICATION NUMBER: 60/242,578
> PRIOR FILING DATE: 2000-10-23
> PRIOR APPLICATION NUMBER: 60/253,625
> PRIOR FILING DATE: 2000-11-27
> PRIOR APPLICATION NUMBER: 60/257,931
> PRIOR FILING DATE: 2000-12-22
> PRIOR APPLICATION NUMBER: 60/269,308
> PRIOR FILING DATE: 2001-02-16
> NUMBER OF SEQ ID NOS: 14110
> SOFTWARE: FastSEQ for Windows Version 4.0
> SEQ ID NO 13862
> LENGTH: 323
> TYPE: PRT
> ORGANISM: Salmonella typhi

```



Db 118 GYLMTIAKVOELLATIDDAYWYNQYANELNMOSHGHGAGTEIVETIKQP--IDYFVAPV 175  
QY 185 GGGGMLAGTAITVKALKSVKVAAPSNADDCYQSKGLKMLPNLYPPETIADGVKSSI 244  
Db 176 STTSGMGRKIKVEHPNQAIIAVD-----AGKSV 206  
QY 245 GLNWTPIIRD-----VDDIFTVTDEIK-----CATOLVWERMKLLIETAGVG 289  
Db 207 IFGDKPINRELPGIGASVRPNEXIRNXSSDPCRXLSICFGLSKTDXLXRHICGSGTGS 266  
QY 290 VAAVLSQHQTQVSPVKNICIVLGGN--VDLTSSITWYKQARPPASQSV 338  
Db 267 IIAAIEQLITSIEGATIVTILDRGDRYLDLVYSDTWL---EKMKSROGV 314

## RESULT 13

US-09-931-457A-32

; Sequence 32, Application US/09931457A  
; Patent No. US20020157132A1

; GENERAL INFORMATION:

; APPLICANT: Falco, S. Carl

; APPLICANT: Allen, Stephen M.

; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes

; FILE REFERENCE: BB1116 US CIP

; CURRENT APPLICATION NUMBER: US/09/931,457A

; CURRENT FILING DATE: 2002-02-22

; PRIOR APPLICATION NUMBER: 09/424,976

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: 60/065,385

; PRIOR FILING DATE: 1997-11-12

; PRIOR APPLICATION NUMBER: 60/049,406

; PRIOR FILING DATE: 1997-06-12

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 32

; TYPE: PRT

; ORGANISM: Citrullus lanatus

US-09-931-457A-32

Query Match 8.5%; Score 148; DB 9; Length 325;  
Best Local Similarity 22.3%; Pred. No. 9.5e-06;  
Matches 64; Conservative 45; Mismatches 98; Indels 80; Gaps 10;

QY 45 KCLFQKTSFKIRGALNAVRSLVDPALERK-----PKAVVTHSSGNHQALTYAAKLE 98  
Db 38 KLEMPECSVKDRIGY----SMISDA-ENKGLITPCESVLIPTSGNTGIGLAFIAAAK 92  
QY 99 GIPAYIVVPTAPDCKKLAIQAYGASIVYCEPS-----DESRENVAK-----140  
Db 93 GYRLIICMPASMSLERRTILRFGAELVLPDARGMKGAVOKAEIKAKTPNSYILQQFE 152  
QY 141 -----RVTEETEGIMVHPNOEPVIAVAGQGTIALEVLNQLVLPVGGGMLAGIAI 195  
Db 153 NPANRKHYYETTG-----PEIWRGSG-----GKIDALVSGIGTGTVTCAGK 194  
QY 196 TVKALKPSVKVVAAPSNADDCYQSKLKGKMLPNLYPPETIADGVK-----SSIGLNTW 249  
Db 195 YLKEQNPNIKLYGVEPES-----AIISSGGPGPHKIQGIGAGFI 234  
QY 250 PIIRD--LVDDIFTVTDEIKCATOLVWERMKLLIETAGVGVAAVL 294  
Db 235 PGVLDVNLDEVIVQSSEESIETAKLLALKEGLLVGISSGAAAAAI 281

## RESULT 14

US-09-905-290A-4

; Sequence 4, Application US/09905290A

; Publication No. US20030087327A1

; GENERAL INFORMATION:

; APPLICANT: Rice, John

; APPLICANT: Lanning, Beth

; APPLICANT: Crawford, John  
; APPLICANT: Nye, Gordon  
; TITLE OF INVENTION: METHODS FOR MEASURING CYSTEINE AND DETERMINING CYSTEINE SYNTHA  
; FILE REFERENCE: Docket No. US20030087327A12090US  
; CURRENT APPLICATION NUMBER: US/09/905,290A  
; CURRENT FILING DATE: 2001-07-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 324  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; US-09-905-290A-4

Query Match 8.5%; Score 147; DB 9; Length 324;

Best Local Similarity 22.3%; Pred. No. 1.2e-05;

Matches 77; Conservative 62; Mismatches 137; Indels 70; Gaps 13;

QY 12 VEKAHINIRDS----IHLTP-VLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVR 66  
Db 1 MEEDRCSIKDDATQLIGNTPMVYLNINIVDGCVAR-TAAKLEMMPECSVVKERIAYGMKD 59  
QY 67 LVPDALERPKA-VVTHSSGNHQALTYAAKLEGIPAYIVVPTAPDCKKLAIQAYGASI 125  
Db 60 AEDKGLITPCGSTLIEATSGNTGIGLAFICAAKGYKVLTMPSSMSLERKIIILLALGAEV 119  
QY 126 VYCEPS-----DESRENVAKRVTEETEGIMVHPNOEPA-----VIAGO 163  
Db 120 HLTPDPSKGVQIIDKAEICSK-----NPDSIMLEQFKNPSNPOTHYTTTGPETWRDSAGE 175  
QY 164 GTIALEVLNQVPLVDALVVPVGGGMLAGIAITVKALKPSVKVYVAAPSNADDCYQSKLK 223  
Db 176 -----VDILVAGVGTGTLSSGSGRELKKNKDFKVVGVPT-----ESAVI 216  
QY 224 GKLMPLNLYPPETIADG--VRKSSIGLNTWPIIRDLVDDIFTVTDEIKCATOLVWERMKLLI 282  
Db 217 SGKPGTHLIQGTGAGLIPDNLDN-----VLDEVIQVTSVEAIEATKALLALKEGLLV 269  
QY 283 EPTAGVGVAAVLSQHQTQVSPVKN-----ICIVLSSGNVDLTSSI 323  
Db 270 GISSGAAAAAIAK-----VAKRPNAGKLLIVIFPSSGGERYLSTSL 310

## RESULT 15

US-09-931-457A-66

; Sequence 66, Application US/09931457A

; Patent No. US20020157132A1

; GENERAL INFORMATION:

; APPLICANT: Falco, S. Carl

; APPLICANT: Allen, Stephen M.

; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes

; FILE REFERENCE: BB1116 US CIP

; CURRENT APPLICATION NUMBER: US/09/931,457A

; CURRENT FILING DATE: 2002-02-22

; PRIOR APPLICATION NUMBER: 09/424,976

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: 60/065,385

; PRIOR FILING DATE: 1997-11-12

; PRIOR APPLICATION NUMBER: 60/049,406

; PRIOR FILING DATE: 1997-06-12

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 66

; LENGTH: 386

; TYPE: PRT

; ORGANISM: Solanum tuberosum

US-09-931-457A-66

Query Match 8.5%; Score 147; DB 9; Length 386;

Best Local Similarity 19.2%; Pred. No. 1.5e-05;

Matches 64; Conservative 60; Mismatches 112; Indels 98; Gaps 10;

QY 8 SFADVEKAHI--NIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVR 65



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 06:07:29 ; Search time 27.0398 seconds  
(without alignments)  
1208.801 Million cell updates/sec

Title: US-09-889-609B-10  
Perfect score: 1735  
Sequence: 1 MCAQYICISFADVEKAHINIR.....SSITWVKQAEPAQSVSV 340  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	802.5	46.3	346	2 T04211	probable threonine
2	605	34.9	326	2 S38061	threonine ammonia-
3	565	32.6	323	2 T41297	threonine ammonia-
4	539	31.1	320	2 E83310	L-serine ammonia-1
5	535	30.8	325	2 T36434	threonine ammonia-
6	532.5	30.7	401	2 D72386	threonine ammonia-
7	517	29.8	346	2 F89921	threonine ammonia-
8	514.5	29.7	317	2 T24337	threonine ammonia-
9	508.5	29.3	323	2 F85390	probable threonine
10	496.5	28.6	324	2 A97687	probable serine/th
11	496.5	28.6	324	2 A29112	threonine dehydrat
12	486.5	28.0	495	2 H84359	threonine dehydrat
13	483	27.8	405	2 D90166	hypothetical prote
14	479.5	27.6	325	2 G87633	threonine dehydrat
15	471	27.1	329	2 A80897	threonine ammonia-
16	470	27.1	343	2 D37778	threonine ammonia-
17	463	26.7	504	2 F83603	threonine dehydrat
18	461.5	26.6	320	2 AG0496	threonine ammonia-
19	457	26.3	329	1 DMECTD	threonine ammonia-
20	457	26.3	329	2 D85973	threonine ammonia-
21	457	26.3	329	2 E91128	threonine ammonia-
22	448.5	25.9	403	2 D81355	threonine ammonia-
23	446.5	25.7	336	2 A31442	threonine dehydrat
24	445.5	25.7	333	2 H71703	probable threonine
25	439	25.3	568	2 E75502	threonine ammonia-
26	437	25.2	512	2 T23166	hypothetical prote
27	424.5	24.5	304	2 A98146	probable threonin
28	424	24.4	334	2 B95894	threonine ammonia-
29	423.5	24.4	600	2 T39516	threonine ammonia-

30.	418.5	24.1	508	2 S77559	threonine ammonia-
31	417	24.0	400	2 A87700	threonine dehydrat
32	416.5	24.0	349	2 C72630	probable threonine
33	403	23.2	515	2 B83479	threonine dehydrat
34	390.5	22.5	576	1 DWBYT	threonine ammonia-
35	385	22.2	514	2 AG0474	threonine ammonia-
36	384	22.1	510	2 E82374	threonine ammonia-
37	379	21.8	422	2 A69644	threonine ammonia-
38	376	21.7	503	2 A12334	threonine ammonia-
39	371	21.4	257	2 A81839	threonine dehydrat
40	368	21.2	592	2 T51712	threonine ammonia-
41	364.5	21.0	422	2 A81694	threonine dehydrat
42	363	20.9	595	2 A38628	threonine ammonia-
43	362	20.9	514	2 AB0924	threonine ammonia-
44	361.5	20.8	422	2 AG1323	threonine dehydrat
45	361	20.8	514	2 B91217	threonine ammonia-

ALIGNMENTS

RESULT 1

T04211  
Probable threonine ammonia-lyase (EC 4.3.1.19) T5C23.70 [similarity] - Arabidopsis th  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 21-Jun-2002  
C:Accession: T04211  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z15261  
A:Accession: T04211  
A:Molecule type: DNA  
A:Residues: 1-346 <BEV>  
A:Cross-references: EMBL:AL049500  
A:Experimental source: cultivar Columbia; BAC clone T5C23  
C:Genetics:  
A:Map position: 4  
A:Introns: 55/2; 99/2; 176/2; 263/2  
A:Note: T5C23.70  
C:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase

Query Match 46.3%; Score 802.5; DB 2; Length 346;  
Best Local Similarity 48.5%; Pred. No. 1.1e-54;  
Matches 163; Conservative 67; Mismatches 83; Indels 23; Gaps 6;

Qy	4	QYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQ-----	50
Db	7	KYAADILSIKEAHDRIKPYIHRTPVLTSESLSISGRSLFFKCECLQKGIERSFYFSVY	66
Qy	51	--KTGSFKIRGALNAVRSLVDPDALEKPKAVVTHSGNGHOALTVAAKLEGIPAYIVVPO	108
Db	67	CKSGAFKFRGACNAVLSL--DA-EQAAGVVVTHSGNHAALSAAKIQGIPAYIVVPK	123
Qy	109	TAPDCKKLAIOAYGASIVCYCPESDESRENKVRKRVTEETEGIMVHPNQEPVIAIGQTTAL	168
Db	124	GAPKCKVDNRYRGKVKVISEATMSRREIAKSVLQETGSLVLIHPYNDGRIISGGQTTAL	183
Qy	169	EVLNQVPLVDALVVPVGGGGLAGTATVTKALKPSKVYAAEPSNADDCYQSKLKGKLM	228
Db	184	ELLEQIQIDIAIVPISGGGSLGVALAAKSIKPSIRIIAEPKGAQAQSKVAGKIT-	242
Qy	229	NLYPETTIADGVKSSIGLNTWPIIRDLDVDDIFTVTEDEIKCATOLVWRMKLLIEPTAGV	288
Db	243	TLPVNTIADGLURASGLDTPVVRDLDDVVVLTLECEIIEAMKMKYELLKVSPEPGAI	302
Qy	289	GVAAVLSQHFQTVSP---EVKNICIVLSGGNVDLTS	321
Db	303	GLAAVLSNSFRN-NPSCRDCKNIGIVLSGGNVDLGS	337

RESULT 2

S38061

[illegible]



C;Keywords: ammonia-lyase; carbon-nitrogen lyase

```
Query Match      29.8%; Score 517; DB 2; Length 346;
Best Local Similarity 37.1%; Pred. No. 1.6e-32;
Matches 118; Conservative 65; Mismatches 129; Indels 6; Gaps 4;

QY 5 YCSFADVEKAHINIRDSIHLPVLTSSILNQ-LTGRNLFKCELFQKTSFKIRGALNA 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 12 HIVSLGDIEEAKASIKPFIIRTPLIKSMYLSQNTKGNVYKLENMQFTSGSFRFGASNK 71

QY 64 VRSIVPDALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPDCKKLAIOYGA 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 72 INHL---SDPKAGKIGASAGNAGQVALTAKLLGLGDATIVMPETAPIAKQNAKGYGA 128

QY 124 SIIVCEPSDESRENKRVTEETEGIMVHPNQEPAVIAGOGTIALEVLNOVPLVDALVVP 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 KVLKGNFNETRLYMEELAKENGMIIVHPYDKFWAGOGTIGLEILDDIWNVNVIYP 188

QY 184 VGGGMLAGIATVTKALPKSVKYAAAPSNADDCYOSKLGKMLPNLYPPETIADGVKSS 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 VGGGGLIAGIATALKSFNPSIHIIIGVQAEVNHGMAESFYKRALTEH-REDSTIADGCDVK 247

QY 244 I-GLNTWPIIRDLDVDFITVTEDEIKCATOLVWERMKLLIEPTAGVGAVALVSOHQFTVS 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 248 VPGEKTYEVVYKHLVDFEILVSEEEIEHAMQDLMQRAKIITEGAGALPTAILSGKIDKKW 307

QY 303 PEVKNICIVLGGNGVDLT 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 LEGKNVVALVSGNGVDLT 325
```

RESULT 8  
T24337  
threonine ammonia-lyase (EC 4.3.1.19) T01H8.2 [similarity] - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jun-2002  
C;Accession: T24337  
R:Lennard, N.  
submitted to the EMBL Data Library, September 1996

A;Reference number: Z19877  
A;Accession: T24337  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-317 <WIL>  
A;Cross-references: EMBL:Z80219; PIDN:CAB02298.1; GSPDB:GN00019; CESP:T01H8.2  
A;Experimental source: clone T01H8  
C;Genetics:  
A;Gene: CESP:T01H8.2  
A;Map position: 1  
A;Introns: 75/3; 131/2; 159/3; 203/1; 259/3; 276/3  
C;Superfamily: threonine dehydratase  
C;Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase; phospho-  
F;56/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted

```
Query Match      29.7%; Score 514.5; DB 2; Length 317;
Best Local Similarity 38.0%; Pred. No. 2.2e-32;
Matches 122; Conservative 61; Mismatches 115; Indels 23; Gaps 8;

QY 7 ISFADVEKAHINIRDSIHLPVLTSSILNQLTGR--NLFFKCELFQKTSFKIRGALNAV 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 ITMEEMRAVQRTAPLHLTKTLVTISNDEKVGNGTHVLFKCEHLQKTSFKARGALNSA 64

QY 65 RSLVPDALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPDCKKLAIOYGA 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 IL----AKEKNAGMIAHSSGNHGQALAWAAQKIGLPCTIVVPKNAPIKIEGMREYNAN 120

QY 125 IVYCEPSDESRENKRVTEETEGIMVHPNQEPAVIAGOGTIALEVLNOV-PLVDALVVP 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 IVFCEPTVTSRESVCADLTKLEYCYCEPYNCSVMINGSHSSVAFLEQVGNEIDSIFLS 180

QY 184 VGGGMLAGIATVTKALPKSVKYAAAPSNADDCYOSKLGKMLPNLYPP-----ETIADG 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 VGGGGLASSVAFILGNLRPDIEVYLWQA-----QKLSNLFLENGIKCPVDTLDTIADG 234
```

```
QY 240 VK-SSIGLNTWPIIRD-LVDDIFTVTEDEIKCATOLVWERMKLLIEPTAGVGAVALVSOH 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 VRVARVGPCEPILKQYCAQKVISVKEDEIKGLKIWTMRKORIEPTAAALAFAGVLYHK 294

QY 298 FQTVSPVKNICIVLGGNGVD 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 ----PAHIKRPVLVILGGNGVD 311

RESULT 9  
F95390  
probable threonine dehydratase [imported] - Sinorhizobium meliloti (strain 1021) mega  
C;Species: Sinorhizobium meliloti  
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 27-Nov-2001  
C;Accession: F95390  
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; B  
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K  
proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli  
A;Reference number: A95262; MUID:21396509; PMID:11481432  
A;Accession: F95390  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-323 <KUR>  
A;Cross-references: GB:AE006469; PIDN:AAK65688.1; PID:g14524179; GSPDB:GN00165  
A;Experimental source: strain 1021, megaplasmid pSymA  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl  
peita, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau  
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,  
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A;Reference number: A96039; MUID:21368234; PMID:11474104  
A;Contents: annotation  
C;Genetics:  
A;Gene: Smal872  
A;Genome: plasmid  
C;Superfamily: threonine dehydratase
```

```
Query Match      29.3%; Score 508.5; DB 2; Length 323;
Best Local Similarity 36.8%; Pred. No. 6.5e-32;
Matches 114; Conservative 62; Mismatches 127; Indels 7; Gaps 4;

QY 11 DVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRLVDP 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 DREARERLKPVHRHTPLLRAEKIERAAGCQLYKLPETLQITGAFKIRGALKALS- 65
```

```
QY 71 ALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPDCKKLAIOYGASIVVCEP 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 SREEIANGIIATSSGNHGQALSLAAKMLGVKVLVLPVTPPKIKIENTKALGADEVILFDG 125

QY 131 SDESRENKRVTEETEGIMVHPNQEPAVIAGOGTIALEVLNOVPLVDALVVPVGGGML 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 DNAARKKVVETAEAGNKYAVIHGFEDPVVMAGOGTIGCEILEDDVDVTIVPLGGGGLI 185

QY 191 AGTATVTKALPKSVKYAAAPSNADDCYOSKLGKMLPNLYPPETIADGVKSSI-GLNTW 249
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 SGATAIKETKPSVRVIGAEPAITPKYFHSRV-NKERTSLPKNTIADGLRISVPCQNPY 244

QY 250 PIRDLVDDIFTVTEDEIKCATOLVWERMKLLIEPTAGVGAVALVSOHQFTVSPVKNIC 309
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 245 PIIEKYVDELVLVDEDEHIITAGMALAKADAKLIAEPAASIGVGALLAGIITDKLDE--KVC 302

QY 310 IVLSGGNGVDL 319
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 303 AVLSSGNGWDL 312
```

RESULT 10  
A97687  
probable serine/threonine dehydratase, degradative (PA2683) [imported] - Agrobacteriu

C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: AF2912  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouellet, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A:Reference number: AF2912; PMID:11743194  
A:Accession: AF2912  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-324 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK88450.1; PID:g15157951; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C4956  
A:Map position: circular chromosome  
C:Superfamily: threonine dehydratase

Query Match 28.6%; Score 496.5; DB 2; Length 324;  
Best Local Similarity 38.7%; Pred. No. 5.6e-31;  
Matches 121; Conservative 48; Mismatches 131; Indels 13; Gaps 5;

QY 12 VEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGALNAVRSILVDPDA 71  
DB 7 IEAARERIGNHVRPLTSPFLDEIAGRKLVKAECLQRTGSFKRGWSAVSGLPADV 66

QY 72 LERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPDCKKLAIAQYASIVYCEPS 131  
DB 67 ---RAKGVIAFSSGNHQAQVLAARLHGIPAVIIMPSPDAPKIKIDNTRAYGAEEVLYDRA 123

QY 132 DESRENKAVRTEETEGIMVHPNQEPNAVIAAGGTIALEVLNQ-----VPLVDALVVPVGG 186  
DB 124 NEDRAIGNRLSSERGLTLIRPYDEPLVIAAGGTAGLEIAEQAGELGIGAAEVL-VPCGG 182

QY 187 GGLAGIAITVKALKPSVKVYAAEPSNADDCVQSKLKGKLMPLNLYPPEIADG-VKSSSTG 245  
DB 183 GGLTSGISLADAKARNYKVRTAEPRFDDVARSIAAGKIERNATTSSGICDAIVTPQPG 242

QY 246 LNTWPIIRDVLDDIFTVTEDEIKCATQLVWERMKLLIEPTAGVGAVALSQHFQTVSPV 305  
DB 243 NITFPIMAGLCCKGIAVSEEEALRAMVLAFNRLKVVIEPFGGAVALAAL---FHGKELES 299

QY 306 KNICIVLSGGNVD 318  
DB 300 ETVIAVASGGNVD 312

RESULT 11  
AF2912  
threonine dehydratase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002  
C:Accession: AF2912  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan; Karp, P.; Romero, P.; Zhang, S.; Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.; E.W.;  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AF2912; PMID:11743193  
A:Accession: AF2912  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-324 <KUR>  
A:Cross-references: GB:AE008688; PIDN:AA43716.1; PID:g17741247; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: illva  
A:Map position: circular chromosome  
C:Superfamily: threonine dehydratase

Query Match 28.6%; Score 496.5; DB 2; Length 324;

Best Local Similarity 38.7%; Pred. No. 5.6e-31;  
Matches 121; Conservative 48; Mismatches 131; Indels 13; Gaps 5;

QY 12 VEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGALNAVRSILVDPDA 71  
DB 7 IEAARERIGNHVRPLTSPFLDEIAGRKLVKAECLQRTGSFKRGWSAVSGLPADV 66

QY 72 LERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPDCKKLAIAQYASIVYCEPS 131  
DB 67 ---RAKGVIAFSSGNHQAQVLAARLHGIPAVIIMPSPDAPKIKIDNTRAYGAEEVLYDRA 123

QY 132 DESRENKAVRTEETEGIMVHPNQEPNAVIAAGGTIALEVLNQ-----VPLVDALVVPVGG 186  
DB 124 NEDRAIGNRLSSERGLTLIRPYDEPLVIAAGGTAGLEIAEQAGELGIGAAEVL-VPCGG 182

QY 187 GGLAGIAITVKALKPSVKVYAAEPSNADDCVQSKLKGKLMPLNLYPPEIADG-VKSSSTG 245  
DB 183 GGLTSGISLADAKARNYKVRTAEPRFDDVARSIAAGKIERNATTSSGICDAIVTPQPG 242

QY 246 LNTWPIIRDVLDDIFTVTEDEIKCATQLVWERMKLLIEPTAGVGAVALSQHFQTVSPV 305  
DB 243 NITFPIMAGLCCKGIAVSEEEALRAMVLAFNRLKVVIEPFGGAVALAAL---FHGKELES 299

QY 306 KNICIVLSGGNVD 318  
DB 300 ETVIAVASGGNVD 312

RESULT 12  
H84359  
threonine dehydratase [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: H84359  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky; Leichauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja; Jung, K.H.; Alam, M.; Freitas, T.; Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: H84359  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-495 <STO>  
A:Cross-references: GB:AE004437; NID:g10581515; PIDN:AAG20244.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: illua

Query Match 28.0%; Score 486.5; DB 2; Length 495;  
Best Local Similarity 36.7%; Pred. No. 6e-30;  
Matches 116; Conservative 59; Mismatches 130; Indels 11; Gaps 7;

QY 7 ISFADVERAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGALNAVRS 66  
DB 87 LSLSDIRARERVSSETARTPLEYSHFTSGLTGADVRPKLECFORTGSFKIRGATNRI 146

QY 67 LVPDALKERPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPDCKKLAIAQYASIV 126  
DB 147 LSAD---QODAGVVTASAGNHAQGVALAASRGVSKVVMPEAPISLKAATKSYGAEVV 203

QY 127 YCEPDESRENKAVRTEETEG-IMVHPNOEPAVIAAGGTIALEVLNQVPLVDALVVPV 185  
DB 204 -LHGADYDDAOAHAHELEAAEGRTVYHAFDVEYIMAGGTGLGLEIAADCTVDTVVVPIG 262

QY 186 GGGMLAGIAITVKALKPSVKVYAAEPSNADDCVQSKLKGKLMPLNLYP-PETTADGVK-SS 243  
DB 263 GGGISGVATALKGELDDVRVIGVQAEGASTVARSLLDKG---POAVDHVDITDIADGIVRQ 320

QY 244 IGLNTPPIIRDVLDDIFTVTEDEIKCATQLVWERMKLLIEPTAGVGAVALSQHFQTVSP 303  
DB 321 VQAQTFPVIQEHVDEVTVVSDDIEATLVLLERKTLVEGAGATALAAVLEKDFEYADG 380



Search completed: June 24, 2003, 06:23:10  
Job time : 28.0398 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:44:58 ; Search time 41.5611 Seconds  
(without alignments)  
1685.614 Million cell updates/sec

Title: US-09-889-609B-10  
Perfect score: 1735  
Sequence: 1 MCAQYCISFADVEKAHINR.....SSITWVQAEKRPASVQSVSV 340

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	802.5	46.3	346	10 Q9T0D1	Q9T0d1 arabidopsis
2	602	34.7	323	16 Q8XPX0	Q8xpx0 ralstonia s
3	550	31.7	388	3 Q8X0J0	Q8x0j0 neurospora
4	539	31.1	320	16 Q910F5	Q910f5 pseudomonas
5	535	30.8	325	16 Q9XAA4	Q9xaa4 streptomyce
6	532.5	30.7	401	16 Q9WVJ1	Q9wvj1 thermotoga
7	517	29.8	346	16 Q9U500	Q9u500 staphylococ
8	514.5	29.7	317	5 Q93968	Q93968 caenorhabdi
9	508.5	29.3	323	16 Q921F8	Q921f8 rhizobium m
10	499	28.8	402	16 Q8XL77	Q8xl77 clostridium
11	497.5	28.7	404	17 Q975Y3	Q975y3 sulfolobus
12	496.5	28.6	324	16 Q8UBW5	Q8ubw5 agrobacteri
13	486.5	28.0	409	16 Q9EWG3	Q9ewg3 streptomyce
14	486.5	28.0	495	17 Q9NNH6	Q9nnh6 halobacteri
15	483	27.8	405	17 Q980P1	Q980p1 sulfolobus
16	480.5	27.7	406	17 Q97CB2	Q97cb2 thermoplasm

#### ALIGNMENTS

#### RESULT 1

Q9T0D1	ID	Q9T0D1	PRELIMINARY;	PRT;	345 AA.
AC	Q9T0D1;				
DT	01-MAY-2000 (TrEMBLrel. 13, Created)				
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)				
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)				
DE	TSC23.70 protein.				
GN	TSC23.70 OR AT4G11640.				
OS	Arabidopsis thaliana (Mouse-ear cross).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,				
RA	Bancroft I., Mewes H.W., Mayer K.F.X., Schueller C.;				
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	EU Arabidopsis sequencing project;				
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,				
RA	Mewes H.W., Lemcke K., Mayer K.F.X.;				
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RA	EU Arabidopsis sequencing project;				
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AL049500; CAB39935.1; .				
DR	EMBL; AL161532; CAB78207.1; .				
DR	HSP; P04968; ITDJ				
DR	InterPro; IPR001926; B6_enzyme_beta.				
DR	InterPro; IPR000634; S/T_dehydrtse.				
DR	Pfam; PF00291; PALP; 1.				
DR	PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.				

Q9A3U7 caulobacter  
Q9HLW2 thermoplasm  
Q8RDT9 fusobacteri  
Q8RVF0 pyrobaculum  
Q8Z3K4 salmonella  
Q92HZ2 rickettsia  
Q9I6G0 pseudomonas  
Q8Z9V3 yersinia pe  
Q9VHF0 drosophila  
Q987A3 rhizobium l  
Q9PP95 campylobact  
Q8UP8 agrobacteri  
Q9ZD93 rickettsia  
Q9RW8 deinococcus  
Q21080 caenorhabdi  
Q95XY8 caenorhabdi  
Q92WC4 rhizobium m  
Q94634 schizosacch  
P73375 synecocyst  
Q9A2D0 caulobacter  
Q9YBV1 aeropyrum p  
Q985M4 rhizobium l  
Q98Q4 rhizobium l  
Q9I418 pseudomonas  
Q9AUG1 oryza sativ  
Q8W314 oryza sativ  
Q9XBI2 bacillus ce  
Q8ZAB4 yersinia pe  
Q9KWL1 vibrio chol

KW Hypothetical protein.  
SQ SEQUENCE 346 AA; 36839 MW; 47C37E1CA52B91F0 CRC64;  
Query Match 46.3%; Score 802.5; DB 10; Length 346;  
Best Local Similarity 48.5%; Pred. No. 3.3e-56;  
Matches 163; Conservative 67; Mismatches 83; Indels 23; Gaps 6;

QY 4 QVCISFADVEKAHINRDSIHLPVLTSSILNQLTGRNLFKCELFQ----- 50  
DB 7 KYAADILSIEKAHRIKPIYHRTPTVLTSSILNQLTGRNLFKCELFQ----- 56  
QY 51 --KTSFVKIRGALNAVRSLVPDALKRPAKVTHSSGNHQAALTYAAKLEGIPAYIVPQ 108  
DB 67 CKSGAFKFRGACNAVLSL--DA-EQAAKGVVTHSSGNHAAALSLAAKIQGIPAYIVPK 123  
QY 109 TAPDCKKLAIOAGSIVCEPSDESRENKRVTEETEGIMVHPNPPAVTAGOQTAL 168  
DB 124 GAPCKVDNIRYGGKVIWSEATMSREIAKSLQETGSLVHPNDGRIISGGOQTAL 183  
QY 169 EVLNQVPLDALVVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYQSKLKGKLM 228  
DB 184 ELLEQIQEIDAIWVPISGGGLISGVALAAKSIKPSIRIIAEPKGAADAAQSKVAGKII- 242  
QY 229 NLYPPTTIADGVKSSIGLNTWPIIRDLDVDTFTVEDEIKCATQLWERMKLLIEPTAGV 288  
DB 243 TLPVTNTIADGLRASGLDTPWVRDLVDVDTFTVEDEIEAKMKCYELKVSPEPGAI 302  
QY 289 GVAAVLSQHFQVSP---EVKNKICIVLSGGNVDLTS 321  
DB 303 GLAAVLSNFRN-NFSCRCKNIGIVLSGGNVDLGS 337

RESULT 2  
Q8XPX0  
ID Q8XPX0 PRELIMINARY; PRT: 323 AA.  
AC Q8XPX0;  
DT 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Probable amino-acid dehydratase protein (EC 4.2.-.-).  
GN RSP1516 OR RS04793.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OG Plasmid megaplasmid.  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OX Ralstonia.  
RN NCBI\_TaxID=305;  
RP [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GM11000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
RA Chandler M., Choise N., Moisan C., Claudel-Robard C., Cunnac S., Demange N.,  
RA Gaspin C., Lavie M., Moisan C., Claudel-Robard C., Saurin W., Schiex T.,  
RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
RA Weissenbach J., Boucher C.A.;  
RA "Genome sequence of the plant pathogen Ralstonia solanacearum."  
RL Nature 415:497-502(2002).  
DR EMBL; AL646085; CAD18667.1; --  
DR InterPro: IPR001926; B6\_enzyme\_beta.  
DR Pfam: PF00291; PALP; 1.  
DR Lyase; Plasmid; Complete proteome.  
KW SEQUENCE 323 AA; 34515 MW; 2175CBF693E04A2D CRC64;  
SQ

Query Match 34.7%; Score 602; DB 16; Length 323;  
Best Local Similarity 40.8%; Pred. No. 3.6e-40;  
Matches 128; Conservative 65; Mismatches 113; Indels 8; Gaps 4;

QY 7 ISFADVEKAHINRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFVKIRGALNAVR 66  
DB 6 ITYDVAHARLQCGVHRTPVLTSSILNQLTGRNLFKCELFQKTSFVKIRGALNAVR 65  
QY 67 LPVDALKRPAKVTHSSGNHQAALTYAAKLEGIPAYIVPPTAPDCKKLAIOAGSIV 126

KW Hypothetical protein.  
SQ SEQUENCE 346 AA; 36839 MW; 47C37E1CA52B91F0 CRC64;  
Query Match 46.3%; Score 802.5; DB 10; Length 346;  
Best Local Similarity 48.5%; Pred. No. 3.3e-56;  
Matches 163; Conservative 67; Mismatches 83; Indels 23; Gaps 6;  
QY 4 QVCISFADVEKAHINRDSIHLPVLTSSILNQLTGRNLFKCELFQ----- 50  
DB 7 KYAADILSIEKHAHRDKPIYHRTPTVLTSSLSISGRSFFKCECLOKIECRSFFSVY 66  
QY 51 --KTSFVKIRGALNAVRSLVPDALEKPKAVVTHSSGNHQAALTYAAKLEGIPAYIVVPQ 108  
DB 67 CKSGAFKFRGACNAVLSL--DA-EQAAKGVVTHSSGNHAAALSLAAKIQGIPAYIVVPK 123  
QY 109 TAPDCKKLAIOAGSIVYCEPSDESRENKRVVTEETEGIMVHPNPPAVTAGOQTAL 168  
DB 124 GAPCKVDNIRYGGKVIWSEATMSREIAKSLQETGSLVHPNDGRIISGGOQTAL 183  
QY 169 EVLNQVPLDALVVPVGGGMLAGIAITVTKALKPSVKVYAAEPSNADDCYQSKLKGKLM 228  
DB 184 ELLEQIQEIDAIVPISGGGLISGVALAAKSIKPSIRIIAEPKGAADAAQSKVAGKII- 242  
QY 229 NLYPPTTIADGVKSSIGLNTWPIIRDLDVDTFTVTEDEIKCATQLWERMKLLIEPTAGV 288  
DB 243 TLPVTNTIADGLRASGLDTPVWRDLVDVDTFTVTEDEIEAKMKCYELTKVSPSGAI 302  
QY 289 GVAAVLSQHFQVSP---EVKNKICIVLSGGNVDLTS 321  
DB 303 GLAAVLSNFRN-NPSCRCKNIGIVLSGGNVDLGS 337  
RESULT 2  
Q8XPX0  
ID Q8XPX0 PRELIMINARY; PRT: 323 AA.  
AC Q8XPX0;  
DT 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Probable amino-acid dehydratase protein (EC 4.2.-.-).  
GN RSP1516 OR RS04793.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OG Plasmid megaplasmid.  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GM11000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
RA Chandler M., Choinse N., Claudel-Robard C., Cunnac S., Demange N.,  
RA Gaspin C., Lavie M., Moisan A., Renard C., Saurin W., Schiex T.,  
RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
RA Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."  
RL Nature 415:497-502(2002).  
DR EMBL; AL646085; CAD18667.1; --  
DR InterPro: IPR001926; B6\_enzyme\_beta.  
DR Pfam: PF00291; PALP; 1.  
KW Lyase; Plasmid; Complete proteome.  
SQ SEQUENCE 323 AA; 34515 MW; 2175CBF693E04A2D CRC64;  
Query Match 34.7%; Score 602; DB 16; Length 323;  
Best Local Similarity 40.8%; Pred. No. 3.6e-40;  
Matches 128; Conservative 65; Mismatches 113; Indels 8; Gaps 4;  
QY 7 ISFADVEKAHINRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFVKIRGALNAVR 66  
DB 6 ITYDDVAHARLQCGVHRTVPLVSTTANATGELFFKCNFMFGAFKFRGNALSQ 65  
QY 67 LPVDALERKPKAVVTHSSGNHQAALTYAAKLEGIPAYIVVPDCKKLAIOAGYASIV 126

```
Db 312 MYSGVEEIKKALRLVYERMKVVVPSAVVGLVAFNFEFRSMVEREGGEGWDLGVVF 371
QY 313 SGGNVDLT'S 321
Db 372 SGGNVELAA 380

RESULT 4
Q910F5
ID Q910F5 PRELIMINARY; PRT; 320 AA.
AC Q910F5;
DT 01-WAR-2001 (TREMBLrel. 16, Created)
DT 01-WAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-WAR-2002 (TREMBLrel. 20, Last annotation update)
DE Probable serine/threonine dehydratase, degradative.
GN PA2683.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Miziochichi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Brinkman F.S.L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Iori S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL: AE004696; AAG06071.1; -.
DR HSSP: P04968; 1TDJ.
DR InterPro: IPR001926; B6_enzyme_beta.
DR Pfam: PF00291; PALP; 1.
KW Complete proteome.
SQ SEQUENCE 320 AA; 33957 MW; C6F97C6259FE8584 CRC64;

Query Match 31.1%; Score 539; DB 16; Length 320;
Best Local Similarity 38.0%; Pred. No. 4e-35;
Matches 119; Conservative 66; Mismatches 120; Indels 8; Gaps 4;

QY 8 SPADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRSL 67
Db 6 TYDDVIAAARAGHANRTPVMSRSLDEELGAEEVFFKCNLQRMGAFFRGAFNALSRF 65
QY 68 VPDALERKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPOTAPDCKKLAIQAYGASIVY 127
Db 66 ---SAEQRAAGVAFSSGNGHAQIALSARLLGIPATIVMPADAPAVKTEATRGYGGQVVL 122
QY 128 CEPSPDESRENKRVTEETEGIMVHPNCEPAVIAQGGTIALEVLNQPLVDALVVPVGG 187
Db 123 YDRYTEDREQIGRLDAQRHGLTLIPPYDHPDLVAGQGTAAKELFEEVGLDPAFFAPLGGG 182
QY 188 GMLAGIATVTKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNLYPPETIADGVKSS-ICL 246
Db 183 GLLSCALAIRALAPACRIYGEVEAGNDGQSRSLSGAIV-HIDTPQTLADGAQTOHLGN 241
QY 247 NTPWPIRLVDVDDIFTVTDEIKCATQLVWERMKLLIETAGVGVAAVLSQHFTQVSPYVK 306
Db 242 LTFPLIQNRVDDILTASDAELVDGMRFLAARMKLLVEPTGCGUGLAAARQKDEL---RGK 298
QY 307 NICIVLSGGNVDL 319
Db 299 RVGILLSGGNVDL 311

RESULT 5
Q9XAA4
ID Q9XAA4 PRELIMINARY; PRT; 325 AA.
AC Q9XAA4;
```

```
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE Putative threonine dehydratase.
GN SC00821 OR SCF43A.11C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K., Harris D.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2)."
RL Nature 417:141-147(2002).
DR EMBL: AL096837; CAB48898.1; -.
DR HSSP: P04968; 1TDJ.
DR InterPro: IPR001926; B6_enzyme_beta.
DR Pfam: PF00291; PALP; 1.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
SQ SEQUENCE 325 AA; 33965 MW; 4E1DE62D340489A7 CRC64;

Query Match 30.8%; Score 535; DB 16; Length 325;
Best Local Similarity 39.6%; Pred. No. 8.5e-35;
Matches 124; Conservative 56; Mismatches 125; Indels 8; Gaps 4;

QY 7 ISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVR 66
Db 8 VTLDVRSAAARKIGVAHRTPVLSRTLDALVGAEVHLKACNQORVGAFFRGAYNAASR 67
QY 67 LVPDALERKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPOTAPDCKKLAIQAYGASIV 126
Db 68 LTPQELAR---GTAAYSSGNGHAQVALARELGTAVIVMPEDAPPSPKRDATAGYAEIV 124
QY 127 YCEPSPDESRENKRVTEETEGIMVHPNCEPAVIAQGGTIALEVLNQPLVDALVVPVGG 186
Db 125 TYDRYTGDRVAVAEALADRGTLTLPPEYEHVPIAGQGTAAKELFEEVGLDPAFFAPVGG 184
QY 187 GMLAGIATVTKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNLYPPETIADG-VKSSITG 245
Db 185 GGLIAGSATAVKALHPGHRVIGVEPEAGDDTTRSLAAGRRV-SVPVPRTIADGQALPTPG 243
QY 246 LNTWPIRLVDVDDIFTVTDEIKCATQLVWERMKLLIETAGVGVAAVLSQHFTQVSPV 305
Db 246 LNTWPIRLVDVDDIFTVTDEIKCATQLVWERMKLLIETAGVGVAAVLSQHFTQVSPV 305
```



RC	STRAIN=1021;
RX	MEDLINE=21396509; PubMed=11481432;
RA	Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA	Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA	Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA	Kalman S., Keating D.H., Palm C., Peck M.C., Surzyski R., Wells D.H.,
RA	Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT	"Nucleotide sequence and predicted functions of the entire
RT	Sinochizobium meliloti pSymA megaplasmid.";
RL	Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR	EMBL: AE007289; AA65688.1; ..
DR	InterPro: IPR001926; B6_enzyme_beta.
DR	InterPro: IPR000634; S/T_dehydrtse.
DR	Pfam: PF00291; PALP; 1.
DR	PROSITE; PS00165; DEHYDRATASE_SER_THR; UNKNOWN.1.
KW	Plasmid; Hypothetical protein; Complete proteome.
SQ	SEQUENCE 323 AA; 34706 MW; 1BF363B67F4D635B CRC64;
Query Match 29.3%; Score 508.5; DB 16; Length 323;	
Best Local Similarity 36.8%; Pred. No. 1.le-32;	
Matches 114; Conservative 62; Mismatches 127; Indels 7; Gaps 4	
QY	11 DVEKAHINIRDSIHLTPVLTSILNQLTGRNLFFKCELFQKTSGFKIRGALNAVRSLYPD 70
Db	9 DIREARLRKPVHRHTLLRAEKTEKAAGCOLYLKPETLOITGAFKIRGALKALS--- 65
QY	71 ALERKPKAVYTHSSGNHQALTAAKLEGIPAYIVVPOTAPDKCKLAIOYGASIVCYEP 130
Db	66 SREIANGIIATSSGNAOGLSYAAKMGLGVKLVLPVTTPKIKENTKALGAEVILFDG 125
QY	131 SDSESRNAVKKRVTTETEGIMVHPNQEPAVIAGQTIALEVLNQVPLDALVPPVGGMGL 190
Db	126 DNAARNKKVYEIABGNKYAVIHGFDPVMVMAQGTFIGCEILLDDVDTVIVPLGGGGLI 185
QY	191 AGTAITVKALKPSVKVYAAPSNAADDVCOSKLKGKMLPNLYPETTIADCVKSSI--GLNTW 249
Db	186 SGIIATAIKETKPSVRVIGAEPALTPKFHSRV-NKERTSLPKNTIADGLRISVPCQNPY 244
QY	250 PIIRDLDVDDLTFTVDEIKCATOLVWERMKLLIEPTACGVAAVLQSOFQTWSPVKNIC 309
Db	245 PILEKYVDEIVLVEDEHIAGMRALAKDAKLIAEPAASIGVALLGIIDVKLDE--KVC 302
QY	310 IVLSGGGNVDL 319
Db	303 AVLSSGGNWDL 312
RESULT 10	
Q8XL77	PRELIMINARY; PRT; 402 AA.
ID	Q8XL77
AC	Q8XL77;
DT	01-MAR-2002 (TREMBLrel. 20, Created)
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Threonine dehydratase.
GN	THD OR CPE1165.
OS	Clostridium perfringens.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC	Clostridiales; Clostridiaceae; Clostridium.
OX	NCBI_TaxID=1502;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=13 / TYPE A;
RX	PubMed=11792842;
RA	Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA	Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT	"Complete genome sequence of Clostridium perfringens, an anaerobic
RT	flesh-eater."
RL	Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR	EMBL: AP003189; BAB80871.1; ..
DR	InterPro: IPR002912; ACT.
DR	InterPro: IPR001926; B6_enzyme_beta.
DR	InterPro: IPR000634; S/T_dehydrtse.



```
Matches 121; Conservative 48; Mismatches 131; Indels 13; Gaps 5;
QY 12 VEKAHINIRDSIHLPVLTSSILNOLTNLFFKCELFQKTSFKIRGALNAVRSLVPPA 71
Db 7 IEAARERIGHAVRTPLLTSPFDELAGRKLKFAECLORTGSFKFRGWSAVSGLPADV 66
QY 72 LERKPAVTHSSGNHGQALTYAAKLEGIPAYIVVPOTAPDCCKLAIOAYGASIVYCEPS 131
Db 67 ---RAKGVIATFSGNHAQGVAAALRHGIPAVIIMPSPDKIKIDNTRAYGAENVLYDRA 123
QY 132 DESRENVAKRVTEETEGIMVHPNOEPVAVIAGQGTIALEVLNQ-----VPLVDALVVPVG 186
Db 124 NEDRDAGNRKSERGLTIRPYDEPLVIAQGTAGLETAEOGAELGIGAAEVL-VPCGG 182
QY 187 GGMLAGIATVTKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNLYPPETIADG-VKSSIG 245
Db 183 GGLTSGISLALDAKARNYKRVTAEPERFDDVARSLAAGKIERNATVSGSICDAIVTPQG 242
QY 246 LNTWPIIRDLVDDITFTVTEDEIKCATQLVWERMKLLIEPTAGVGVAIVLSQHFTQVSP 305
Db 243 NITFPIMAGLCGRGIAVSEELRAMVLAFNRLKVVIEPGGAVAAAAAL---FHGKELES 299
QY 306 KNICIVLSGGNVD 318
Db 300 ETVIAVSGGNVD 312

RESULT 13
Q9EWG3 PRELIMINARY; PRT; 409 AA.
AC Q9EWG3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative amino acid deaminase.
GN 25CK31.22 OR SCO4962.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Harris D.E., Ouail M.A., Kieser H.,
RA Thomson N.R., James K.D., Chatter K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
```

```
RL Nature 417:141-147(2002).
DR EMBL; AL451182; CAD18697.1; -
DR AL512667; CAD30948.1; -
DR HSSP; P04968; 1TDJ
DR InterPro; IPR001926; B6_enzyme_beta.
DR InterPro; IPR000634; S/T_dehydrtse.
DR Pfam; PF00291; PALP; 1.
DR TIGRFAMS; TIGR01127; ilva.1cterm; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 409 AA; 42279 MW; 9F21A40D3CEA9B9 CRC64;

Query Match 28.0%; Score 486.5; DB 16; Length 409;
Best Local Similarity 38.2%; Pred. No. 9.1e-31;
Matches 120; Conservative 54; Mismatches 129; Indels 11; Gaps 6;

QY 7 ISFADVEKAHINIRDSIHLPVLTSSILNOLTNLFFKCELFQKTSFKIRGALNAVRSLV 66
Db 12 VTLDVDRGAKMLSGVARVTAMEGRHLSQVLSGSPVHLKCEQLORTGSFKIRGAYVRIAG 71
QY 67 LVPDALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPOTAPDCCKLAIOAYGASI - 125
Db 72 LLP---EQRAGVVAASAGNHAQGVAAALRHGIPAVIIMPSPDKIKIDNTRAYGAENV 128
QY 126 VYCEPSDESRENVAKRVTEETEGIMVHPNOEPVAVIAGQGTIALEVLNOVPLVDALV 185
Db 129 LHGVVVDETLA-AAQEVARTGAVFIHPDPDVIAGQGTIVGLEILEQCPEVTVVVGIG 187
QY 186 GGMLAGIATVTKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNLYPPETIADGVK-SSI 244
Db 188 GGGLAAGIAVAVKALRPDVRIVGVQAEAA-AAYPPSLAAGRPAVAVONPATMADGIKVG 246
QY 245 GLNTWPIIRDLVDDITFTVTEDEIKCATQLVWERMKLLIEPTAGVGVAIVLSQHFTQV 304
Db 247 GDVPFGIVGELVDEVTVSDEALSTALLCLERAKLVVEPAGASPAVAAALLSEPDAGEP - 305
QY 305 KNICIVLSGGNVD 318
Db 306 ---VWAVLSGGNVD 316

RESULT 14
Q9HNNH6 PRELIMINARY; PRT; 495 AA.
AC Q9HNNH6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Threonine dehydratase.
GN ILUA OR VNG2100G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005099; AAG20244.1; -
DR HSSP; P04968; 1TDJ
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001926; B6_enzyme_beta.
DR InterPro; IPR000634; S/T_dehydrtse.
DR Pfam; PF01842; ACT; 1.
```

